Journal of Research in Science and Engineering (JRSE)
ISSN: 1656-1996 Volume-6, Issue-12, December 2024
DOI: 10.53469/jrse.2024.06(12).06

Analysis of the Factors Influencing Carbon Trading Prices in Shenzhen

Song Hao

Guangxi Normal University, Guilin, Guangxi, China

Abstract: ***China's carbon emissions trading market is in the ascendant. As the country with the largest carbon emissions in the world, China is actively taking measures to deal with global climate change. China has put forward the long-term goal of reaching the peak of carbon emissions by 2030, achieving carbon neutralization by 2060 and shouldering important international social responsibilities. The construction and development of carbon emissions trading market is a key link in achieving the goal of carbon peak and carbon neutralization. The research on the influencing factors of China's regional carbon emissions trading price and the development of regional green finance can provide decision support for achieving the goal of carbon peak and carbon neutralization in China. Starting from the four latitudes of domestic and foreign economies, domestic and foreign energy prices, international carbon prices, and international major currencies and RMB exchange rate, this paper selects 13 independent variables to explore their impact on the dependent variable Shenzhen carbon trading average price. First of all, this paper uses lasso for regression modeling, and selects the λ that minimizes the mse into the model. Five features of the model are screened out. Among the remaining characteristics, the CSI 300 index has a greater positive effect on the Shenzhen carbon trading price, while the SSE 50 has a greater reverse effect on the Shenzhen carbon trading price. Secondly, in order to avoid using lasso to compress the model coefficients too small, this paper uses the elastic network model to fit the data, the fitting result is actually the ridge regression model, and the prediction mean square error of the test set is greater than that of lasso modeling. Finally, considering all the characteristics comprehensively, this paper uses the random forest algorithm to model the data, and the characteristics that make a great contribution to the carbon trading price in Shenzhen are EU emission quota and Guangdong carbon trading price. and the mean square error of the test set modeled by random forest algorithm is the smallest of the three models, and the prediction effect is the best.***

Keywords: Carbon trading price, Lasso, Elastic network, Random forest**.**

1. Introduction

1.1 Research Background and Significance

Carbon dioxide is a type of carbon oxide, typically a colorless, odorless gas at normal temperature and pressure. It is a key component of the atmosphere and a common greenhouse gas. Greenhouse gases absorb long-wave radiation reflected from the Earth's surface and re-emit it, warming the planet, a phenomenon known as the "greenhouse effect." As the global economy develops, the impact of the greenhouse effect has become increasingly evident, making energy conservation, emissions reduction, and carbon emission control a crucial issue for both China and the world. By 2021, China's annual carbon dioxide emissions had reached 10.52 billion tons, ranking first in global total emissions. Compared to the 78.58 million tons in the early years of the country's founding, China's annual carbon dioxide emissions have increased more than 133 times. Energy conservation and emissions reduction have thus become a significant issue for China's development. Since the "11th Five-Year Plan" period, China has actively pursued energy conservation and emissions reduction, gradually achieving its reduction targets. In 2005, China set clear energy-saving and consumption-reduction targets in the "11th Five-Year Plan," and in September 2020, during the 75th session of the United Nations General Assembly, China formally announced its carbon neutrality and peak carbon dioxide emissions goals [1]. The Chinese government stated that it would implement more stringent policies and measures to reach peak carbon emissions before 2030, reduce per capita carbon emissions by 60%-65% by 2030, and achieve carbon neutrality by 2060 [2].

Carbon trading refers to the trading of greenhouse gas emission rights. The Kyoto Protocol, established in December 1997 in Kyoto, Japan, set requirements to reduce six types of greenhouse gases, with carbon dioxide being the most significant. Unlike traditional energy-saving and emissions-reduction measures, the Kyoto Protocol introduced a market-based mechanism and defined carbon dioxide emission rights as commodities. Only those possessing carbon emission rights can emit the corresponding amount of carbon dioxide, guiding industrial production toward energy-saving, low-carbon, and efficient development, thus reducing greenhouse gas emissions and mitigating the global warming caused by the greenhouse effect.

China's energy-saving and emissions-reduction policies have gradually shifted from a government-mandated enforcement mechanism to a market-oriented mechanism based on carbon trading rights. In 2011, the Chinese government approved pilot carbon trading programs in Beijing, Shanghai, Tianjin, Chongqing, Shenzhen, Guangdong, and Hubei to explore market-driven emissions reduction mechanisms. In November 2015, China's President proposed the establishment of a nationwide carbon emissions trading market at the Paris Climate Conference, a step with long-term significance for addressing China's climate change challenges. In October 2016, the State Council issued the "13th Five-Year Plan for Controlling Greenhouse Gas Emissions," which outlined plans to begin establishing a national carbon trading market in 2017. On December 19, 2017, China officially launched its carbon market, initiating carbon trading in the power sector and gradually expanding it to other industries and products [3].

To establish a nationwide carbon trading market-based emissions reduction mechanism, it is essential to learn from the experiences of the carbon trading pilot programs. From the near-term goal of reaching peak carbon emissions by 2030 to

the long-term goal of achieving carbon neutrality by 2060, China has only 30 years to achieve carbon neutrality, which is half the buffer time compared to advanced countries that typically need around 50 years. However, the internal development of the carbon trading market is still in its early stages [4], and its price regulation mechanisms are not yet fully developed. The fluctuation of carbon trading prices is influenced by many factors, and such fluctuations can affect enterprises' participation in carbon trading, which in turn hinders the fair allocation of carbon trading rights and ultimately impacts the sustainable development of the domestic carbon market. Therefore, exploring the factors influencing carbon trading prices is of great importance.

China's carbon trading market-based mechanism is still in its early stages, and the current approved carbon trading pilot programs play a crucial role in the future establishment of a national carbon trading market. Research on carbon trading prices and their influencing factors is of great theoretical and practical significance for stabilizing the carbon trading market and guiding the development of a market-oriented mechanism in a positive direction.

Firstly, since China's carbon trading market-based mechanism is still in its infancy, there is a scarcity of related research, and even fewer studies focus on the factors influencing carbon trading prices. Therefore, this paper aims to fill this gap. Moreover, previous domestic research in this area has mostly been limited to analyzing domestic factors. This study not only incorporates domestic factors that may affect carbon trading prices but also considers international energy and economic factors, providing a more comprehensive understanding of the variables influencing carbon trading prices.

Investigating the factors related to carbon trading prices will help promote the healthy development of the market-oriented carbon trading mechanism, build corporate confidence in the carbon trading market, and provide important guidance for the future realization of a nationwide carbon trading market.

1.2 Related Work

Both domestic and international studies have conducted in-depth discussions on the international pricing mechanisms of carbon emission rights and the related influencing factors, with certain results already being established. D. Sokalakis et al. (2009) considered the impact of various factors on the EU ETS EUA prices [5]. Bunn and Fuzzi primarily analyzed the role of energy in determining EU ETS EUA prices and concluded that natural gas significantly influences carbon trading prices [6]. Alberola and Chevalier (2008) found that the EU carbon allowance prices are not only related to the prices of energy sources like natural gas but also to changes in temperature [7]. Regarding research methods, Kanen (2006) studied the variation in carbon trading prices from the perspective of supply and demand balance in the carbon trading market [8]. Some scholars have also used big data and data analysis models, such as linear regression models, VAR models, and autoregressive models, to analyze carbon prices [9].

1.3 Research Content and Methodology

This study explores the impact of 13 independent variables on the dependent variable, the average carbon trading price in Shenzhen, from four dimensions: domestic and international economies, domestic and international energy prices, international carbon prices, and the exchange rate between major international currencies and the Chinese yuan. Firstly, the study employs the Lasso regression method for model building, selecting the λ that minimizes the mean squared error (MSE). The model resulted in the exclusion of five features, with the remaining features showing that the CSI 300 index has a significant positive effect on Shenzhen's carbon trading price, while the SSE 50 index has a notable negative effect. Secondly, to prevent Lasso from shrinking the model coefficients too much, the study uses the Elastic Net model to fit the data. The fitting results essentially resemble a ridge regression model. However, the prediction MSE for the test set from the Elastic Net model is higher than that from the Lasso model. Finally, taking all features into account, the study uses the Random Forest algorithm for modeling. The model reveals that the European Union Emission Allowance and Guangdong carbon trading prices are significant contributors to Shenzhen's carbon trading prices. Moreover, the prediction MSE for the test set from the Random Forest model is the smallest among the three models, yielding the best predictive performance.

This study applies statistical methods, specifically regression analysis, and utilizes R language for modeling and data analysis. The main techniques used include comparative analysis and panel data analysis. In the comparative analysis approach, the study selects 13 independent variables related to carbon trading prices for analysis. The Lasso regression method is employed to filter variables, and models are constructed using Lasso, an Elastic Net model combining ridge regression and Lasso, and the Random Forest algorithm. These models are compared to identify the most effective one. For panel data analysis, the study focuses on evaluating model performance by examining the monthly average carbon trading prices in Shenzhen from 2017 to February 2022, along with the 13 related variables. Different models are used to fit the data, and the best-performing model is selected, ensuring that the identified factors influencing carbon trading prices are more reliable.

2. Methodology

2.1 Prerequisite Knowledge

2.1.1 Overfitting

Overfitting occurs when a model fits the training data too well, capturing not only the underlying patterns but also the noise or random fluctuations in the data. While this leads to excellent performance on the training set, the model's ability to generalize to new, unseen data is poor. In other words, the model has learned details specific to the training data that don't hold in broader contexts, resulting in poor predictive performance on new datasets.

2.1.2 Structural Risk Minimization

The average loss of a model on the training set is called empirical risk (Remp).

$$R_{emp}(f) = \frac{1}{N}\sum_{i=1}^{N} L\big(y_i, f(x_i)\big) \quad (1)$$

A loss function measures the difference between the predicted and actual values after training. The most common types of loss functions include:

(1) 0-1 Loss Function

$$L\big(y, f(x)\big) = I\{y = f(x)\} \quad (2)$$

(2) Absolute Loss Function

$$L\big(y, f(x)\big) = |y - f(x)| \quad (3)$$

(3) Squared Loss Function

$$L\big(y, f(x)\big) = \big(y - f(x)\big)^2 \quad (4)$$

(4) Logarithmic Loss Function

$$L\big(y, P(Y|X)\big) = -logP(Y|X) \quad (5)$$

During the process of minimizing empirical risk, overfitting is likely to occur. To avoid this issue, we introduce a penalty term on top of the empirical risk.

$$R_{srm}(f) = \frac{1}{N}\sum_{i=1}^{N} L\big(y_i, f(x_i)\big) + \lambda J(f) \quad (6)$$

$J(f)$: Model complexity: The more complex the model, the larger the $J(f)$

λ: A coefficient that measures the trade-off between empirical risk and model complexity, where $\lambda \geq 0$

Therefore, the Structural Risk Minimization we aim to achieve is:

$$min(f) = \frac{1}{N}\sum_{i=1}^{N} L\big(y_i, f(x_i)\big) \quad (7)$$

2.1.3 Regularization

Regularization is the implementation of the strategy of Structural Risk Minimization. The added penalty term is also known as the regularization term. Therefore, the strategy of adding a regularization term to the empirical risk is called regularization. The penalty term is proportional to the model complexity, meaning the higher the model complexity, the higher the value of the penalty term.

In regression problems, the penalty term is typically referred to as the norm. The L_p norm is defined as:

$$\|w\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}} \quad (8)$$

2.2 Ridge Regression

Ridge regression improves upon the least squares estimation by adding the L2 norm $\big|\big|w\big|\big|_2$ to the residual sum of squares, which minimizes the residual sum of squares while avoiding overly large coefficients. Assuming the independent variables are standardized and the dependent variable is centered, the objective function is:

$$Q = (y - X\beta)^T (y - X\beta) + \lambda \sum_{j=1}^{n} {\beta_j}^2 \quad (9)$$

where $\lambda(\geq 0)$ is the ridge parameter.

For the objective function with the penalty term, we can always find a λ such that $(X^T X + \lambda I)$ is invertible. Therefore, we can take the partial derivative with respect to β, set it equal to zero, and obtain the ridge regression estimate of β:

$$\beta^T = (X^T X + \lambda I)^{-1} X^T y \quad (10)$$

Clearly, when $\lambda = 0$, the above equation is equivalent to the ordinary least squares estimate of the multivariate linear regression coefficients β. When $\lambda \to \infty$, $\beta^T \to 0$.

The value of λ can be determined through cross-validation, by using the Cp criterion, or by examining the ridge trace plot; however, the ridge trace plot can be relatively subjective. When multicollinearity exists in the data, ridge regression often yields better results. Ridge regression can mitigate the effects of multicollinearity to some extent. Compared to ordinary multivariate linear regression, although ridge regression sacrifices the unbiasedness of the parameter estimates and introduces some bias, it effectively reduces the variance of the estimated parameters.

2.3 Lasso

Lasso (Least Absolute Shrinkage and Selection Operator) is an improvement upon least squares estimation by adding the L1 norm $\|w\|_1$. Compared to ordinary multivariate linear regression, like ridge regression, Lasso provides biased estimates, but it also makes the model more stable and prevents overfitting. The key difference between Lasso and ridge regression is that Lasso can produce sparse solutions. For independent variables that have little impact on y, Lasso will quickly shrink their coefficients to zero, whereas ridge regression only drives the coefficients toward zero but does not exactly make them zero.

As shown in Figure 1, this behavior is illustrated with two independent variables as an example:

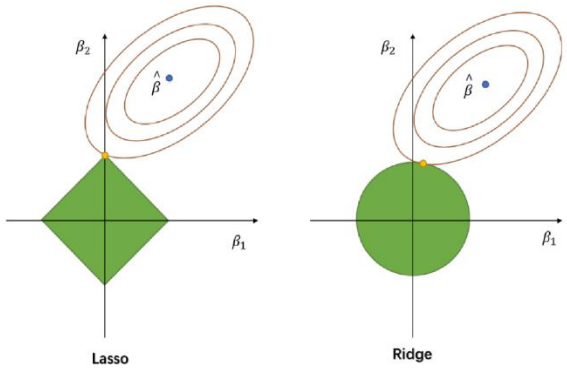

Figure 1: Illustration of Lasso and Ridge Regression Solutions

In the figure, the blue points represent the regression coefficients obtained from the least squares estimate, while the orange elliptical contour lines indicate the contours for different values of λ. The green region represents the constraint space for both Ridge and Lasso regression. The orange dots mark the points where the contour lines (representing the L1 or L2 penalty) are tangent to the constraint space, which correspond to the estimated values of Ridge and Lasso regression.

It is easy to observe that Lasso can shrink some coefficients to exactly zero, whereas Ridge regression cannot. In

high-dimensional cases, Lasso can easily shrink many parameters to zero, leading to a sparse solution where only the most relevant variables remain, while Ridge regression generally keeps all parameters non-zero but shrinks them towards smaller values.

2.4 Elastic Net

The Elastic Net is based on least squares estimation and simultaneously incorporates both the L1 norm and L2 norm. The objective function for Elastic Net is:

$$J_E(w) = \frac{1}{n}\|y - Xw\|^2 + \lambda[p\|w\|_1 + (1-p)\|w\|^2] \quad (11)$$

where:

$$\hat{\beta}_{Elastic} = arg\min_{\beta \in R^d}\left\{\|Y - X\beta\|^2 + \lambda\left[p\sum_{j=1}^{d}|\beta_j| + (1-p)\sum_{j=1}^{d}|\beta_j|^2\right]\right\} \quad (12)$$

2.5 Random Forest

2.5.1 Regression Tree

A regression tree is similar to an additive model, as both represent a compromise between linear models and fully non-parametric methods. Suppose X and Y are the input and output variables, respectively, and Y is a continuous variable. Given a training dataset $D = \{(x\_1, y\_1), (x\_2, y\_2), \dots, (x\_N, y\_N)\}$, a regression tree corresponds to a partition of the input space (i.e., the feature space) and assigns an output value within each partition. Assuming the input space is divided into M regions $R_1, R_2, \dots, R_M$, and each region R_m has a fixed output value c_m, the regression tree model can be expressed as:

$$f(x) = \sum_{m=1}^{M} c_m I(x \in R_m) \quad (13)$$

When the partition of the input space is determined, the regression tree's prediction error for the training data can be represented by the squared error:

$$\sum_{x_i \in R_m}[y_i - f(x_i)]^2 \quad (14)$$

The optimal output value for each region is found by minimizing the squared error criterion. It is easy to see that the optimal value of c_m for the region R_m is the mean of the output values y_i corresponding to all input instances x_i in R_m, that is:

$$\widehat{c_m} = \frac{1}{|R_m|}\sum_{x_i \in R_m} y_i \quad (15)$$

Therefore, the problem is how to partition the input space. We can use a heuristic method, and the steps are as follows:

a) Choose the $j-th$ variable $x(j)$ and its value s as the splitting variable and splitting point, and define two regions:

$$R_{1(j,s)} = \{x \mid x(j) \le s\},\ R_{2(j,s)} = \{x \mid x(j) > s\} \quad (16)$$

Then, find the optimal splitting variable j and the optimal splitting point s. Specifically, solve:

$$\min_{j,s}\left[\min_{c_1}\sum_{x_i \in R_1(j,s)}(y_i - c_1)^2 + \min_{c_2}\sum_{x_i \in R_2(j,s)}(y_i - c_2)^2\right] \quad (17)$$

For a fixed input variable j, the optimal splitting point s can be found:

$$c_1 = \text{ave}\big(y_i \mid x_i \in R_1(j,s)\big),\ c_2 = \text{ave}\big(y_i \mid x_i \in R_2(j,s)\big) \quad (18)$$

By iterating over all input variables, the optimal splitting variable j can be found, resulting in a pair (j, s).

b) Use the selected pair (j, s) to partition the region and determine the corresponding output values:

$$R_{1(j,s)} = \{x \mid x(j) \le s\},\ R_{2(j,s)} = \{x \mid x(j) > s\} \quad (19)$$

$$c_m = \frac{1}{N_m}\sum_{x_i \in R_m} y_i,\ x \in R_m\ (m = 1,2) \quad (20)$$

c) Continue applying steps a and b to the two sub-regions until the stopping criteria are met.

d) Partition the input space into M regions $R_1, R_2, \dots, R_M$, and generate the regression tree:

$$f(x) = \sum_{m=1}^{M} \widehat{c_m}\, I(x \in R_m) \quad (21)$$

This type of regression tree is typically called a Least Squares Regression Tree.

2.5.2 Classification Tree

Trees can be used for various types of response data. We can extend tree methods to other types of response variables by fitting an appropriate zero model in each partition. For example, we can extend trees to binomial, multinomial, Poisson, and survival data by using a deviance instead of RSS (Residual Sum of Squares). Classification trees are similar to regression trees, except that the sum of squared residuals is no longer the appropriate criterion for choosing split points.

We can measure the purity of a node in several ways. Let n_{ik} be the number of observations of type k in terminal node i, and p_{ik} be the proportion of type k in node i. Let D_i be the metric at node i, so the total metric is $\sum_i D_i$. There are several choices for D_i:

a) Deviance: $D_i = -2\sum_k n_{ik}\log p_{ik}$
b) Entropy: $D_i = -2\sum_k p_{ik}\log p_{ik}$
c) Gini index: $D_i = 1 - \sum_k p_{ik}^2$

All of these metrics have the same property: they are minimized when all members of the node are of the same type.

2.5.3 Tree Pruning

For determining the optimal size of a tree, one strategy is to keep partitioning until the reduction in the total cost (the Residual Sum of Squares, RSS) is less than a threshold ϵ. However, setting ϵ reasonably can be difficult. Additionally, this strategy might stop early, as each expansion of the tree may not always reduce RSS. Moreover, RSS tends to underestimate the predictive power of the tree, which is a common phenomenon in most models.

To obtain a regression tree with better predictive power, a common method is cross-validation (CV). For a given tree, leave out one case, rebuild the tree with the remaining cases, and use the tree to predict the left-out case. For regression

trees, the optimization criterion is:

$$\sum_{j=1}^{n}\left[y_j - \widehat{f_{(j)}}(x_j)\right]^2 \quad (22)$$

where $\widehat{f_{(j)}}(x_j)$ represents the output of the tree for input x_j when case j is not in the tree construction. For other types of trees, a different criterion is used. For example, for classification problems, deviance can be used as the optimization criterion.

For trees, leave-one-out cross-validation can be expensive, so k-fold cross-validation is often used. The data is randomly split into k roughly equal parts, with $(k-1)$ parts used to train the regression tree and the remaining part used to predict. In addition to being more computationally efficient than leave-one-out cross-validation, it may even be more effective. However, its drawback is that the data partitioning is random, so repeating the method gives different numerical results. Moreover, if we consider all subsets of a large tree, there can be many possible trees. This makes k-fold cross-validation too expensive.

Similar to shrinkage methods, it can be useful to consider a balance of cost and complexity when pruning a regression tree, as this can reduce the set of trees to those that are valuable. We define the cost-complexity function of the tree as:

$$CC(tree) = \sum_i RSS_i + \lambda \times |T| \quad (23)$$

where $|T|$ is the number of nodes. If λ is large, the tree will be smaller to minimize this cost, and vice versa. We can train a large tree to determine the optimal tree for any given size and then prune it. Given a tree of size n, we can determine the best tree of size $n-1$ by considering possible combinations of adjacent nodes. We choose the tree that increases the fit by the least amount. This strategy is similar to backward elimination in linear regression variable selection, and it can be proven to generate the best tree of the given size.

2.5.4 Random forest

Random Forests are essentially a combination of classification decision trees, where randomization is applied both in terms of the variables used and the data samples. This results in the generation of many classification trees, and the results from these trees are then aggregated. Random Forests improve predictive accuracy without significantly increasing computation time. They are insensitive to multicollinearity and can perform well in making predictions even with a large number of explanatory variables (up to thousands).

Random Forests use the bootstrap method with repeated sampling. From the original training sample set, k samples are randomly selected with replacement to form a new training set. Then, k decision tree models are learned from these samples. Finally, the results of the k decision trees are aggregated to form the final prediction. For classification problems, the aggregation method is majority voting, and for regression problems, it is simple averaging. In the process of generating decision trees, if there are M input variables, F variables are randomly selected for each split at each internal node. The splitting criterion at each internal node is based on the best split of these F variables. The value of F is a constant throughout the construction of the random forest model.

3. Shenzhen Carbon Trading Price Influencing Factors Analysis

3.1 Selection of Variables

First, this study selects appropriate dependent and independent variables: The dependent variable is the carbon trading price of the Shenzhen Emissions Trading Center, which is one of the seven most representative carbon exchanges in China. From the perspectives of domestic and international economics, energy, and foreign exchange, this study selects 13 independent variables to analyze the factors influencing China's carbon trading prices.

3.1.1 Dependent Variable

The Shenzhen Carbon Trading Center, established in 2010, is a market-oriented comprehensive platform for emissions reduction and trading services. It is the first Chinese exchange to introduce foreign investment and has led the country in carbon financial innovation, with its indicators in spot trading volume reaching over 100 million and 1 billion yuan. The turnover rate has ranked first nationwide for six consecutive years, making it the most influential carbon exchange in the domestic green and low-carbon environmental sector. To ensure data accuracy, this study focuses on the Shenzhen carbon trading market price (SZA), which, despite its late start and large trading volume, serves as the object of analysis for the study of the factors influencing carbon prices in China's regional markets.

3.1.2 Independent Variables

In order to comprehensively include the factors influencing domestic regional carbon prices in the research model, this study considers 13 indicators based on four aspects: domestic and international economic trends, domestic and international energy prices, international carbon prices, and exchange rates.

Table 1: Variable Symbols and Descriptions of Variables

Symbols	Explanation of Symbols
SZA	Shenzhen Carbon Trading Price
EUA	Futures Settlement Price (Continuous): EU Emission Allowances (EUA)
SHEA	Shanghai Carbon Trading Price
BEA	Beijing Carbon Trading Price
GDEA	Guangdong Carbon Trading Price
HS300	CSI 300 Index
SZ50	SSE 50 Index
WTI	Futures Settlement Price (Continuous): WTI Crude Oil
DCOAL	Futures Closing Price (Active Contract): Coking Coal
NYMEX	Futures Closing Price (Continuous): NYMEX Natural Gas
CNYtoUSD	USD to RMB Exchange Rate
CNYtoEURO	EUR to RMB Exchange Rate
SP500	US: S&P 500 Index
DAX30	Frankfurt DAX Index

3.2 Descriptive Statistics

The data selected in this study ranges from January 2017 to February 2022, sourced from Kaggle. The daily data for the same indicator are averaged to obtain monthly data, and missing values are filled using the average of the preceding and following values.

3.2.1 Shenzhen Carbon Trading Price Trend

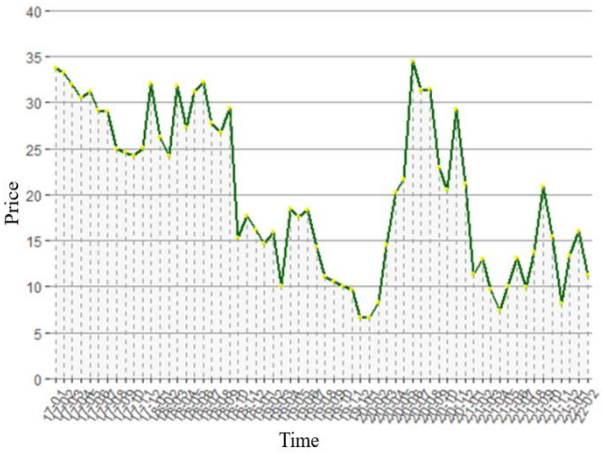

Figure 2: Shenzhen Carbon Trading Price Trend

From the line chart of carbon trading prices in Shenzhen from January 2017 to June 2021, it can be seen that China's carbon emission allowance prices are unstable and do not exhibit seasonal or cyclical characteristics.

To provide a clearer display of the relationships between the independent variables and the dependent variable, a heatmap was created:

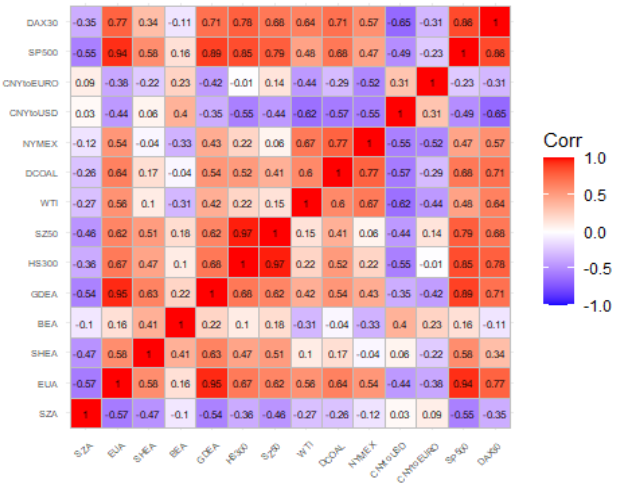

Figure 3: Heatmap of Variables

From the heatmap, we can observe that most of the 13 independent variables are negatively correlated with the dependent variable, Shenzhen Carbon Trading Price (SZA). The only variables that show a positive correlation are the exchange rates of USD to RMB (CNYtoUSD) and EUR to RMB (CNYtoEURO), but the correlation is weak. The strongest correlations are observed with the EU Emission Allowances (EUA), Shanghai Carbon Trading Price (SHEA), and the S&P 500 Index (SP500).

3.3 Model Construction

3.3.1 Lasso Regression Model

Using the Lasso method, as λ increases, some variables are eliminated. If λ is too large, too many variables will be removed, and if λ is too small, the sparsity of the solution may not be achieved. In this case, we use the mean squared error (MSE) of the model as the criterion. To find an appropriate λ, we let the program automatically generate 1000 λ values for fitting. The coefficient variation of each regression variable under these 1000 λ values is shown in the following plot:

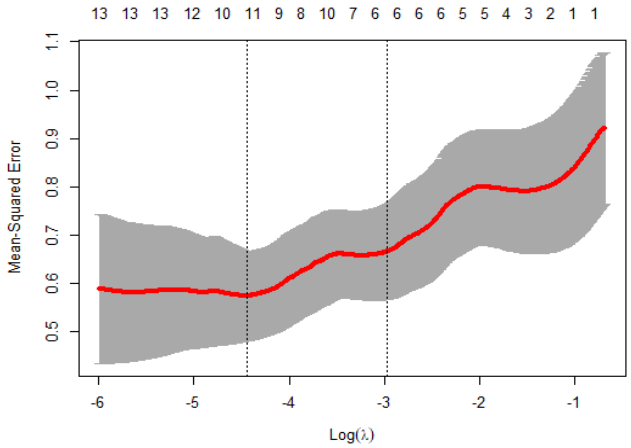

Figure 4: Lasso Plot

The two vertical dashed lines in the plot correspond to the λ value at which the MSE is minimized and the λ value obtained by adding one standard deviation to the λ value at the minimum MSE. This λ value represents a more sparse solution, where more variables are shrunk to zero without significantly increasing the mean squared error. Here, we select the λ value that minimizes the model's MSE and substitute it into the model to obtain the sparse solution and the corresponding parameters as follows:

Table 2: Lasso Regression Coefficients for Independent Variables

Variable	Regression Coefficient
BEA	0.21150578
GDEA	-0.03381008
HS300	2.01272733
SZ50	-1.93900319
DCOAL	-0.03045600
CNYtoUSD	-0.28782294
CNYtoEURO	0.30055745
SP500	-0.70996402

Lasso shrinks the five variables—EU Emission Allowances (EUA), Shanghai Carbon Trading Price (SHEA), Crude Oil (WTI), Natural Gas (NYMEX), and Frankfurt DAX Index (DAX30)—to zero, indicating that, from the perspective of lasso, these variables are insignificant. Using this regression model, the mean squared error between the predicted values and actual values for the test set is 50.179833.

3.3.2 Elastic Net Regression Model

To find the optimal λ and α, we use the expand. grid function to combine different λ and α values and then use the train function to identify the best combination. The optimal model parameters are as follows:

Table 3: Elastic Net Regression Variable Coefficients

Variable	Regression Coefficient
EUA	-0.08602606
BEA	0.18426177
GDEA	-0.01955947
HS300	1.35121597
SZ50	-1.40069592
WTI	-0.05351832
CNYtoUSD	-0.32748007
CNYtoEURO	0.23874202
SP500	-0.52258987

Using this regression model, we obtained a mean squared error (MSE) of 58.604037 for the predicted values compared to the actual values in the test set. This is smaller than the MSE for Lasso, which is not difficult to explain. Compared to Lasso, the Elastic Net model includes the Beijing carbon trading price (BEA), Guangdong carbon trading price (GDEA), the HS300 index, the SZ50 index, the USD/RMB exchange rate (CNYtoUSD), and the EUR/RMB exchange rate (CNYtoEURO). However, the Elastic Net model does not include coal (DCOAL) but instead includes the EU Emissions Allowances (EUA), crude oil (WTI), and the S&P 500 index (SP500). By appropriately increasing the model's complexity, the Elastic Net model achieved a smaller prediction error.

3.3.3 Random forest Model

In the establishment of the Random Forest model, an important parameter is the number of features randomly selected for each tree, which is reflected in the value of mtry. We set up a loop where the number of trees in the random forest is fixed at 500, and mtry varies from 1 to 13, representing the number of features from 1 to all 13 features being randomly selected for each tree. We calculate the residuals of the random forest model on the training set under these conditions. The value of mtry that minimizes the residuals on the training set is found to be 8, meaning the optimal value is when 8 features are randomly selected for each tree.

Next, we use this optimal mtry value of 8 and input it into the Random Forest model, setting ntree to 1000. The curve of the mean squared error (MSE) change with the increasing number of trees is shown below:

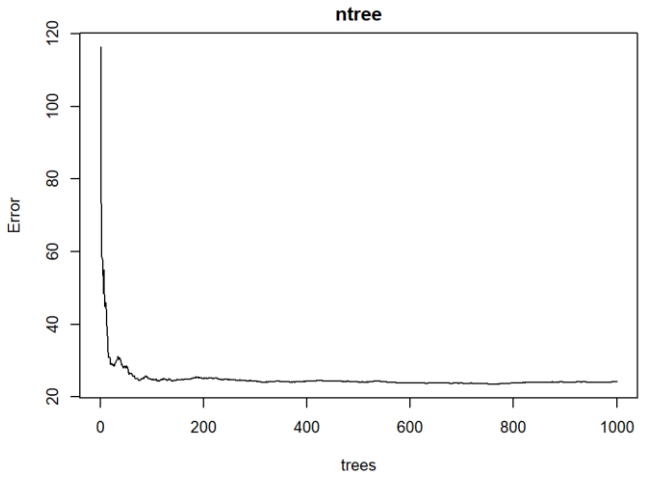

Figure 5: The relationship between the number of trees and the error in the Random Forest model.

From the figure, we can see that the error is minimized when the number of trees in the random forest is approximately 770. With the optimal parameters mtry = 8 and ntree = 770, we apply them to the final random forest model. We then use this model to predict the test set and obtain a prediction Mean Squared Error (MSE) of 27.73. This shows a significant improvement in prediction error compared to the models built using Lasso and Elastic Net. Finally, we examine the importance of each feature in the random forest model, which is shown in the following figure:

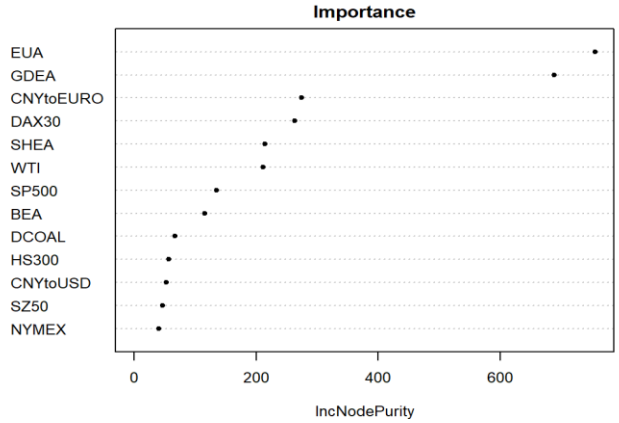

Figure 6: Feature Importance in Random Forest

From the figure, we can roughly categorize all the features into three groups. The most important features are the EU Emissions Allowances (EUA) and Guangdong Carbon Trading Price (GDEA). The second group includes the Euro to RMB exchange rate (CNYtoEURO), the Frankfurt DAX Index (DAX30), Shanghai Carbon Trading Price (SHEA), and Crude Oil (WTI). The remaining features have the weakest impact on the dependent variable, Shenzhen Carbon Trading Price (SZA).

4. Conclusion

In this study, we first used Lasso and Elastic Net for modeling. The modeling results show that the test set prediction mean squared error (MSE) using Elastic Net is larger than the one using Lasso alone. This is because Elastic Net performed poorly on this dataset, essentially behaving like Ridge Regression. To account for all the features, we then used the Random Forest algorithm for modeling. The Random Forest model, considering all features, achieved the lowest MSE on the test set.Through the above models, we can conclude that the factors that significantly impact Shenzhen's carbon trading price are: EU Emissions Allowances (EUA) and Guangdong Carbon Trading Price (GDEA), both of which are domestic and international energy factors, as well as the influence of the exchange rates between the RMB and international currencies.

References

[1] Guo Hongxu, Huang Ying, Fan Junchen, Wang Zhiwei. Research on the CO2 Emission Reduction Effect of Energy-saving and Emission-reduction Measures for Urban Residents' Travel. *Environmental Protection*, 2021, 49(10): 60-66.

[2] Zhou Jialong. The Status and Development Strategies of High Energy Consumption and High Emission Industries in Liaoning Province. *Environmental Protection and Circular Economy*, 2021, 41(08): 100-102+107.

[3] Liu Lei, Zhang Yongqiang. Research on the Carbon Tax System Based on the Carbon Emission Trading Market. *Taxation Research*, 2019, 2019(02): 46-52.

[4] Zhang Lili, Zhang Houmei. Guangyuan City: Accelerating Green Transformation and Striving to Achieve Carbon Peak and Carbon Neutrality. *Chinese Environment*, 2021, 2021(06): 59-61.

[5] Montagnoli A, Vries F P D. Carbon Trading Thickness and Market Efficiency. *Energy Economics*, 2010, 32(6): 1331-1336.
[6] Zou Yasheng, Wei Wei. Research on the Factors Affecting the Spot Price of Certified Emission Reductions (CER). *Financial Research*, 2013, (10): 142-153.
[7] Zhu Bangzhu, Wang Ping, Wei Yiming. Multi-Scale Analysis of the Factors Affecting Carbon Market Prices Based on EMD. *Economic Dynamics*, 2012, (6): 92-97.
[8] Wang Wenjun, Bai Lin. Research on the Factors Affecting the EU Carbon Allowance Price. *Journal of Yunnan Normal University (Philosophy and Social Sciences Edition)*, 2013, 45(04): 135-143.
[9] Wang Junfeng, Zhang Jingwen, Liu Xin. Research on the Correlation Mechanism of Carbon Allowance Prices in the Carbon Emission Trading Market: Based on Econometric Models.