ISSN: 2710-0170

Research on Price Influencing Factors and Machine Learning Pricing Prediction in Carbon Trading Market

--An Empirical Analysis Based on Seven Major Carbon Trading Markets in China

Xunyang Li *, Ruizi Liu

School of Finance, Anhui University of Finance and Economics, Bengbu, China

*1747418404@qq.com

Abstract

As an important tool for addressing climate change and realizing the transition to a low-carbon economy, the carbon trading market has received increasing attention in the study of its price formation mechanism and influencing factors. However, most of the existing studies focus on theoretical discussions and lack empirical analysis of the seven major carbon trading markets in China. In this paper, differential econometric models and machine learning algorithms (including decision trees, random forests, and Xgboost) are used to analyze carbon trading market data between 2014 and 2023 to explore the impact of macroeconomic factors on carbon trading prices. The results of the study show that there is a significant negative correlation between macroeconomic factors and carbon trading prices, and the machine learning model outperforms the traditional linear regression model in predicting carbon trading prices. The research in this paper provides an important reference for improving the efficiency of the carbon trading market and formulating related policies.

Keywords

carbon trading; econometric difference modeling; machine learning

1. Introduction

The theoretical focus of the carbon trading market is the "Peguy Tax" and "Coase's Law", and the logic of the carbon trading market to internalize the externality cost of carbon emissions mainly lies in the design of the total amount of control and trading system, and the hidden economics behind carbon trading are The hidden economics behind carbon trading is the "green premium" (Sun Xia et al., 2024)[3] . This system design reflects the simple justice concept of "who pollutes, who governs; who develops, who protects". As an important tool for coping with climate change and realizing the transformation of low-carbon economy, carbon trading market has received more and more attention in recent years. With the global emphasis on carbon emission control, countries have established and improved the carbon trading market to promote the efficient allocation of resources and technological innovation. As the world's largest emitter of carbon dioxide, China's promotion of low-carbon economic transformation is of great significance to the realization of sustainable development. The primary goal of low-carbon economic transformation is to reduce carbon emissions, and the impact of the traditional energy consumption structure makes carbon emissions from energy consumption the most important source of emissions at present.

Existing studies have shown that an efficient carbon market can effectively reflect the true cost of carbon emissions and thus guide enterprises to adopt low-carbon technologies. Despite the many advantages of the carbon trading market, there are still many limitations in practice, such

ISSN: 2710-0170

as restricted market development, low trading frequency, and low percentage of CER offsets. These problems may be related to the insufficient incentives and limited coverage of the carbon emissions trading market. Therefore, it is of great theoretical and practical significance to deeply analyze the factors affecting carbon trading prices, especially the role of macroeconomic factors.

The innovation of this paper is to use the machine learning approach to deeply analyze the pricing problem in the carbon trading market, which overcomes the limitations of the traditional prediction model that is greatly affected by the data frequency and other factors, and thus improves the accuracy of the prediction results. Although the existing literature affirms the positive role of carbon trading in energy saving and emission reduction, the imperfections of the market lead to problems such as inefficiency, unfair competition and manipulation. Therefore, this paper will select data from seven major carbon trading markets in China during the period from 2014 to 2023 to analyze the impact of macroeconomic factors on carbon trading prices, aiming to improve the efficiency of the carbon trading market and provide a reference for the formulation of related policies. The structure of this paper is organized as follows: the second part is the literature review, the third part introduces the research design and methodology, the fourth part shows the empirical results, and the fifth part is the conclusion and recommendation.

2. Literature review

The carbon trading market is an important tool for combating climate change and realizing low-carbon economic transformation. As countries pay attention to carbon emission control, how to improve the efficiency of the carbon trading market, optimize resource allocation, and promote technological innovation has become a focus of attention for academics and policy makers. This paper reviews related literature to discuss the importance of improving the efficiency of the carbon trading market, the factors affecting the carbon trading price, the limitations of the market and the corresponding solutions.

An efficient carbon market can effectively reflect the true cost of carbon emissions, thus guiding enterprises to adopt low-carbon technologies and realize the efficient allocation of resources. Xu and Liu (2024) point out that, as a large CO2 emitting country, it is of great significance to promote the transition to a low-carbon economy in order to implement the concept of green and low-carbon development. The primary goal of low-carbon economic transformation is to reduce carbon emissions, and affected by the traditional energy consumption structure, carbon emissions from energy consumption are the most important source of emissions at present, and also the focus of carbon emission reduction in the future[5] . At the same time, Wuzheni et al. (2024) emphasized that, from the perspective of implied carbon emissions, the differences in the implementation of carbon trading policies will increase the uneven distribution of carbon reduction responsibility between pilot and non-pilot regions, which enriches the relevant conclusions of the current research on the impact of carbon trading pilot policies on implied carbon emissions between regions[4] . Despite the many advantages of the carbon trading market, there are still many limitations in actual operation. Luo Liangwen et al. (2024) pointed out that the domestic carbon emissions trading market still faces some challenges, such as limited market development, inability to be applied on a large scale, low trading frequency, low percentage of CER offsets, complicated offset procedures, and difficulty in mobilizing market enthusiasm. This may be related to the lack of incentives in the carbon emissions trading market, or it may be due to the limited coverage of carbon trading, resulting in poor policy implementation[2] . Sun Xia and Liang Hongzhi (2024) emphasize that the phenomenon of hidden carbon emissions transfer not only affects the emission reduction effect, but also may lead to unfair competition between regions. In addition, the market mechanism is not perfect

ISSN: 2710-0170

enough and the lack of regulatory measures makes the market vulnerable to manipulation, similar to the carbon trading market manipulation problem mentioned in Zhaoetal.'s (2023) study[7][14] .

The formation of carbon trading price is the result of a combination of factors. According to Gao Kai et al. (2024), policy regulation, market supply and demand relations and various external economic environments affect the volatility of carbon trading prices[1] . Specifically, policy uncertainty often leads to fluctuations in market participants' expectations of future prices, exacerbating market instability, and Yinetal.(2019) points out that market transparency and information symmetry also play an important role in price formation, and information asymmetry may lead to manipulative behavior of certain participants[12] . In addition, further research by Fangetal.(2023) shows that there is a significant interaction between breakthroughs in science and technology innovation and carbon trading policies, and that a favorable policy environment can promote technological progress, which in turn affects carbon trading prices[8][9] . Therefore, an in-depth understanding of the dynamics and complexity of the carbon market is the basis for effective policy formulation.

In summary, the existing literature affirms the positive role of carbon trading in energy saving and emission reduction, but at the same time, the imperfection of the market leads to its inefficiency, unfair competition, and even manipulation, etc. Therefore, this paper starts from the market, and selects the data of the seven major carbon trading markets in China from 2014 to 2023, and analyzes the impact of macro indicators on their carbon trading, which not only improves the efficiency of the carbon trading market, but also can provide direction for the formulation of relevant policies.

3. Research design

3.1. Research methodology

3.1.1. Differential econometric modeling

In this chapter, in order to simplify the research process and develop an effective econometric model, we will construct a linear model for the relationship between financial asset returns and macroeconomic factors in each pilot market. Specifically, we use the Differenced Econometric Model, which improves the robustness and predictive power of the model by differentiating the time-series data in order to eliminate non-stationarity and seasonal fluctuations.

$$\Delta y_t = \alpha + \gamma(L)\Delta y_t + \beta\Delta x_t + \varepsilon_t$$

Δy_t denotes the difference of the logarithm of carbon price, L denotes the lag factor, and X denotes the vector of explanatory variables. In constructing the linear model, the variables selected mainly contain the trading prices of the seven major trading markets, various indices of the domestic stock market, major energy sources and the U.S. stock index.

The basic idea of differential econometric modeling is to eliminate trend components and cyclical fluctuations in the data by calculating the differential values of the variables (i.e., the difference between the current value and the value at the previous moment), making the data more consistent with the assumption of smoothness. This process not only helps to improve the explanatory power of the model, but also effectively reduces the impact of autocorrelation on the regression results. With the difference treatment, the model can more accurately capture the dynamic relationship between carbon trading prices and macroeconomic factors.

3.1.2. Decision trees

In this paper, Zhou Liang (2022) defines three machine learning models, in order, as base modeling[6] . The three selected decision tree based machine learning algorithms include CART

ISSN: 2710-0170

decision tree algorithm, Random Forest based on Bagging integration algorithm and Xgboost method based on Boosting integration algorithm.CART decision tree performs attribute selection by comparing the Gini index of each attribute.

$$\hat{y}_i^{(t)} = \sum_{k=1}^{t} f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i), f_k \in F, i \in n$$

Where$\hat{y}_i^{(t)}$ is the value of t decision trees summed over the sample as a prediction, f_1 is the tth regression tree, and F is the set space of all regression trees.

3.1.3. Random forests

Random Forest (RF) is an integrated learning method where the model uses voting to combine predictions, designating the most voted category as the final output. Random Forests can handle very large amounts of data, while at the same time, it has an error rate for most learning tasks that is almost at the same level as any other method and has a much smaller tendency to overfitting.

Building a loss function based on a decision tree:

$$L^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t)}\right) + \sum_{k=1}^{t} \Omega(f_k)$$

where l represents the degree of deviation between the predicted value$\hat{y}_i^{(t)}$ and the true valuey_I and Ω represents the complexity of each tree, calculated as:

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda||\omega||^2$$

where T is the number of leaf nodes, w is the weight of the leaf nodes, and γ and λ are the regularity coefficients.

3.1.4. Gradient lifter

Xgboost is proposed as a novel integrated learning method for gradient boosting based on the traditional GBDT model. Compared with the traditional GBDT, Xgboost seeks the optimal solution by adding a regular term to the loss function using a second-order Taylor expansion, which avoids overfitting to a certain extent.

Based on the first two formulas, Xgboost Taylor-expanded the loss function to a quadratic term at$y_i^{(t-1)}$ and transformed the optimization objective into the problem of solving the minimum of a univariate quadratic function given the decision tree structure. Then the greedy algorithm continuously tries to partition the leaf nodes and compares the gain of the objective function before and after the partition until the optimal model is obtained.

3.1.5. Performance evaluation

In order to assess the prediction performance of the above four models, the mean absolute error (MAE) and root mean square error (RMSE) are used as the evaluation metrics for the model goodness of fit[6] , which are calculated using the following formulas:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i|$$

ISSN: 2710-0170

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}((\hat{y}_i - y_i)^2)}$$

N represents the sample size,$\widehat{y_I}$ andy_I represent the predicted and observed values of the sample, respectively, and the smaller MAE and RMSE represent the better fit of the model.

3.2. Variable setting

The data in this paper comes from Wind database, and the monthly data from 2014 to 2023 are used in order to show the changing trend of carbon trading market comprehensively. Referring to the study of Zeng et al. (2023), this study sets the average price of carbon trading transaction (yuan/ton) as the explanatory variable[13] , and divides the explanatory variables into the following categories:

Stock index category: including SSE index, Shenzhen index, energy index, raw material index, industrial index, bond yield;

Energy price category: covering raw coal price (yuan/ton), steel price (yuan/ton), ore and iron ore price (yuan/ton), liquefied natural gas price (yuan/ton), gasoline price (yuan/ton), and crude oil price (yuan/ton), of which the price of raw coal selects the price of Dongsheng in Inner Mongolia, and the gasoline selects the price of No. 95 gasoline[11] ;

U.S. stock index category: includes the NASDAQ and the Dow Jones. The abbreviations of the relevant variables are shown in the table below:

Table1 Variable Definitions

explanatory variable	Average carbon trading price	CTP
explanatory variable		
stock market index	SSE (Shanghai Stock Exchange)	SCI
	SSE (Shenzhen Stock Exchange) index	SZCI
	Energy index	EI
	Raw material index	MI
	industrial index	II
	bond proceeds	BY
energy price	raw coal price	RCP
	steel prices	SP
	Price of Ore and Iron	IOP
	Liquefied natural gas prices	LNG P
	petrol price	GP
	crude oil price	COP
US stock market index	NASDAQ (stock exchange)	NCI
	Dow Jones industrial average (Wall street stock market index)	DJIA

ISSN: 2710-0170

3.3. Data descriptive statistics

Table1 Sample Observation Time and Data Status

Market	Starting time	end time	Total number of observations	Number of non-zero trading days	Percentage of non-zero trading days
Shenzhen	2014/1/1	2023/12/31	3652	2027	55.50%
Shanghai	2014/1/1	2023/12/31	3652	1498	41.02%
Beijing	2014/1/1	2023/12/31	3652	1468	40.20%
Guangdong	2014/1/1	2023/12/31	3652	2142	58.65%
Tianjin	2014/1/1	2023/12/31	3652	897	24.56%
Hubei	2014/1/1	2023/12/31	3652	2238	61.28%
Chongqing	2014/1/1	2023/12/31	3652	843	23.08%

This paper collects the trading activities of China's seven major carbon trading markets during the period from January 1, 2014 to December 31, 2023, reflecting the significant differences in the number of non-zero trading days and their proportions among the markets. The number of non-zero trading days in the Shenzhen market is 2,027, accounting for 55.50% of the total number of observed days, showing a high level of trading activity, while the number of non-zero trading days in the Tianjin and Chongqing markets is only 897 and 843, respectively, with ratios as low as 24.56% and 23.08%, suggesting that they have a low level of trading activity. These differences may be closely related to factors such as market mechanism, policy environment and participant base. In particular, the proportion of non-zero trading days in the Guangdong and Hubei markets was 58.65% and 61.28% respectively, further emphasizing the diversity of market activity.

Table2 Results of descriptive statistics for data indicators

Column	Obs	Mean	Std
CTP	840	32.80	23.46
SCI	840	3102.76	433.87
SZCI	840	10946.09	2084.65
EI	840	994.99	263.93
MI	840	3245.03	678.46
II	840	3977.54	856.90
BY	840	127.44	14.48
RCP	840	438.05	266.85
SP	840	3600.50	886.67
IOP	840	477.64	386.77
LNGP	840	4266.87	1479.23
GP	840	3502.51	4052.17
COP	840	63.95	19.79
NCI	840	8604.20	3568.57
DJIA	840	25370.13	6578.51

The above table shows the statistical data of 840 observations, covering the average price of carbon trading transactions and a number of economic indicators. The mean value of the average price of carbon trading transactions is RMB 32.80 per ton, with a standard deviation of 23.46, indicating that the market price fluctuates significantly and may be affected by changes in policy and supply and demand. In terms of stock indices, the average value of the SSE index is 3102.76, with a standard deviation of 433.87, showing relatively stable market performance; while the average value of the Shenzhen index is 10946.09, with a standard deviation as high

ISSN: 2710-0170

as 2084.65, reflecting greater volatility. The average value of the Energy Index is 994.99, with a standard deviation of 263.93, showing the price volatility of the energy market.

In terms of raw material prices, the average value of raw coal prices was RMB 438.05 per ton, with a standard deviation of 266.85, indicating high price volatility. The average value of steel and mineral iron ore prices was RMB 3,600.50 and RMB 477.64, with a standard deviation of 886.67 and 386.77 respectively, reflecting the volatility of the raw material market. The average LNG and gasoline prices were $4,266.87 and $3,502.51, with standard deviations of 1,479.23 and 4,052.17, respectively, showing the high volatility of energy prices. The mean values of the NASDAQ and Dow Jones indices were 8604.20 and 25370.13, with standard deviations of 3568.57 and 6578.51, respectively, indicating significant volatility in the United States stock market.

4. Empirical results

4.1. Differential regression results

The following table demonstrates the results of regression analysis between various types of financial assets and macroeconomic factors in the seven major carbon trading markets in China. By comparing the regression coefficients, t-values, standard errors (Std) and coefficients of determination (R^2) of different markets, the dysfunctions of the carbon trading market in terms of pricing mechanism can be observed.

Table3 Differential Econometric Model Regression Results

	Shenzhen subprovincial city in Guangdong, special economic zone close Hong Kong				Beijing, capital of People's Republic of China				hillsides			
	ratio	t	Std	R^2	ratio	t	Std	R^2	ratio	t	Std	R^2
SCI	0.000479	1.04	0.00046	0.0010	-0.000020	-0.09	0.00021	0.0214	-0.000359	-1.73	0.00021	0.0680
SZCI	-0.000042	-0.48	0.00009	0.0022	0.000074	1.82	0.00004	0.0329	0.000087	2.22	0.00004	0.0154
EI	0.000579	1.36	0.00043	0.0251	-0.000100	-0.50	0.00020	0.0215	-0.000069	-0.36	0.00019	0.0234
MI	-0.000051	-0.20	0.00026	0.0005	0.000007	0.06	0.00012	0.0273	-0.000019	-0.16	0.00012	0.0226
II	-0.000132	-0.54	0.00024	0.0032	-0.000199	-1.77	0.00011	0.0257	-0.000090	-0.83	0.00011	0.0440
BY	0.039299	2.41	0.01632	0.0861	0.017923	2.36	0.00759	0.0112	0.009829	1.34	0.00734	0.0000
RCP	0.000241	0.70	0.00034	0.0264	0.000487	3.04	0.00016	0.0704	0.000141	0.91	0.00015	0.0560
SP	-0.000143	-1.11	0.00013	0.0001	0.000034	0.57	0.00006	0.0022	-0.000041	-0.70	0.00006	0.0207
IOP	-0.000008	-0.11	0.00007	0.0004	0.000072	2.29	0.00003	0.0085	0.000051	1.67	0.00003	0.0019
LNGP	0.000010	0.31	0.00003	0.0059	-0.000037	-2.46	0.00001	0.0163	0.000007	0.51	0.00001	0.0240
GP	0.004076	0.82	0.00499	0.0054	0.002224	0.96	0.00232	0.0235	0.010335	4.60	0.00225	0.2295
COP	0.000004	0.16	0.00002	0.0048	0.000006	0.61	0.00001	0.0012	0.000017	1.61	0.00001	0.0649
NCI	-0.000017	-0.22	0.00008	0.0021	-0.000148	-4.12	0.00004	0.0313	-0.000014	-0.41	0.00003	0.0156
DJIA	0.000006	0.16	0.00004	0.0347	0.000048	2.50	0.00002	0.0033	0.000009	0.47	0.00002	0.0233

Continued from table 3:

Hubei				Shanghai				Tianjin				Chongqing			
ratio	t	Std	R^2	ratio	t	Std	R^2	ratio	t	Std	R^2	ratio	t	Std	R^2
-0.000134	-0.99	0.00014	0.0565	0.000298	1.05	0.00028	0.0086	-0.000279	-1.22	0.00023	0.0000	-0.001230	-1.78	0.00069	0.0005
0.000010	0.39	0.00003	0.0353	-0.000001	-0.03	0.00005	0.0125	0.000028	0.64	0.00004	0.0185	0.000134	1.15	0.00012	0.0112
-0.000270	-2.17	0.00012	0.0056	0.000460	1.40	0.00033	0.1539	-0.000526	-1.84	0.00029	0.0375	0.000924	1.79	0.00052	0.0001
0.000103	1.39	0.00007	0.0009	-0.000252	-1.28	0.00020	0.0059	0.000033	0.27	0.00012	0.0031	0.000069	0.20	0.00035	0.0016
-0.000078	-1.13	0.00007	0.0040	-0.000060	-0.38	0.00016	0.0216	0.000034	0.29	0.00012	0.0031	0.000215	0.58	0.00037	0.0016
0.020609	5.17	0.00398	0.2773	-0.012840	-1.31	0.00978	0.0035	-0.004508	-0.48	0.00933	0.0227	-0.029033	-1.14	0.02555	0.0001
-0.000024	-0.24	0.00010	0.0880	0.000602	2.09	0.00029	0.0808	0.000121	0.70	0.00017	0.0047	0.000042	0.10	0.00044	0.0004
-0.000037	-1.01	0.00004	0.0022	0.000079	1.02	0.00008	0.1789	-0.000158	-1.77	0.00009	0.0136	-0.000302	-1.73	0.00017	0.0143
0.000046	2.32	0.00002	0.0157	-0.000018	-0.42	0.00004	0.0009	0.000050	1.29	0.00004	0.0042	0.000099	1.10	0.00009	0.0150
0.000013	1.46	0.00001	0.0002	0.000007	0.37	0.00002	0.0623	0.000028	1.25	0.00002	0.0955	-0.000019	-0.41	0.00005	0.0191
0.005120	3.58	0.00143	0.0615	-0.002045	-0.68	0.00300	0.0401	0.005060	2.11	0.00240	0.0237	-0.001444	-0.21	0.00673	0.0176
0.000003	0.48	0.00001	0.0477	0.000014	1.06	0.00001	0.0082	0.000016	1.01	0.00002	0.1042	0.000100	2.04	0.00005	0.0542
-0.000028	-1.25	0.00002	0.0563	-0.000007	-0.16	0.00005	0.0158	0.000044	1.16	0.00004	0.1009	0.000101	0.97	0.00010	0.0096
0.000006	0.53	0.00001	0.1227	0.000033	1.32	0.00002	0.0850	0.000003	0.13	0.00002	0.0325	-0.000049	-0.95	0.00005	0.0048

The significance of the regression coefficients reflects the degree of influence of each macroeconomic factor on the carbon trading market price. In the Shenzhen market, the regression coefficient of the SSE index is 0.000479 and the t-value is 1.038654, indicating that its influence on carbon trading prices is insignificant; in other markets, such as Beijing and Shanghai, the correlation coefficients are negative and the t-values are lower than the thresholds, which further suggests that the relationship between prices and macroeconomic factors in these markets is weak. This phenomenon may reflect the insufficient sensitivity of the carbon trading market to macroeconomic fluctuations, leading to the failure of the market pricing mechanism.

The coefficient of determination (R^2) is generally low in all markets, especially in the Tianjin and Chongqing markets, where the value of R^2 is close to zero, indicating that the model has limited explanatory power for carbon trading prices. This low explanatory power may be due to the insufficient knowledge of market participants about carbon trading, information

ISSN: 2710-0170

asymmetry and imperfect market structure, which results in the market price not reflecting the real supply and demand relationship effectively.

The regression results for some markets show a negative correlation between macroeconomic factors and carbon trading prices. For example, the raw materials index and the energy index show negative coefficients in several markets, suggesting that fluctuations in these factors may have a dampening effect on carbon trading prices. This phenomenon may be closely related to changes in the policy environment of the carbon market, the behavioral patterns of market participants, and the external economic environment, further exacerbating market pricing failures.

The current pricing mechanism in China's carbon market has significant failures, which are mainly reflected in the insignificant influence of macroeconomic factors on market prices, the insufficient explanatory capacity of the market, and the mismatch between prices and the relationship between supply and demand. This phenomenon suggests that policy makers and market participants need to strengthen the regulation and guidance of the carbon market, and improve market transparency and participants' cognitive level, in order to promote the healthy development and effective pricing of the carbon trading market.

4.2. Machine learning regression results

In order to further study and investigate the effectiveness of carbon pricing in the seven carbon trading markets, the econometric prediction model is re-established by combining the influencing factors of carbon trading price. In order to improve the prediction accuracy of the model, and overcome the influence of different frequency data such as year, month, day, and so on on the prediction accuracy, this paper introduces three kinds of machine-learning prediction models, namely, the Decision Tree, the Random Forest of , and the Gradient Booster, on the basis of traditional econometric prediction models. The MAE and RMSE predicted by them are compared with the results predicted by the traditional models to further put forward scientific and effective suggestions on carbon pricing in the seven carbon trading markets.

By comparing the mean absolute error (MAE) and root mean square error (RMSE) of the traditional linear regression model with the three machine learning models (decision tree, random forest, and gradient enhancer) in different regions (Shenzhen, Shanghai, Beijing, Guangdong, Tianjin, Hubei, and Chongqing), it can be seen that the machine learning model significantly outperforms the traditional linear regression model in terms of its ability to predict carbon trading prices.

Table4 Comparison of Decision Tree Prediction Model with Traditional Linear Regression

Markrt	traditional linear regression		decision tree		Accuracy Improvement	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
Shenzhen	11.4712	14.5619	7.9663	11.5283	30.55%	20.83%
Shanghai	11.5224	13.5827	3.8280	7.9028	66.78%	41.82%
Beijing	10.8836	15.8954	8.4643	12.8534	22.23%	19.14%
Guangdong	11.5224	13.5827	4.5404	9.7162	60.59%	28.47%
Tianjin	6.6147	8.7882	7.2256	12.3121	-9.24%	-40.10%
Hubei	5.5342	8.0484	2.8737	5.3237	48.07%	33.85%
Chongqing	8.9913	10.9072	8.7650	14.0732	2.52%	-29.03%

ISSN: 2710-0170

Continued from Table 4

random forest		Accuracy Improvement		Gradient lifts		Accuracy Improvement	
MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
5.86	8.30	48.87%	43.03%	5.75	7.86	49.89%	46.01%
4.22	6.93	63.37%	49.00%	4.38	7.08	62.01%	47.87%
8.01	13.15	26.39%	17.29%	8.08	13.18	25.78%	17.08%
2.95	6.45	74.37%	52.48%	2.91	4.97	74.73%	63.42%
6.08	8.24	8.13%	6.27%	5.67	8.59	14.28%	2.24%
3.41	5.88	38.45%	27.00%	3.68	6.23	33.49%	22.64%
8.37	11.47	6.91%	-5.18%	8.10	11.47	9.89%	-5.20%

In terms of overall values, all machine learning models exhibit lower error values in the MAE and RMSE metrics, indicating better predictive capabilities. For example, in the Shenzhen region, the MAE of the decision tree is 7.97 and the RMSE is 11.53, which is a significant reduction compared to the MAE (11.47) and RMSE (14.56) of the traditional linear regression, with an improvement in accuracy of 30.55% and 20.83%. Similar trends are verified in other regions, especially in Shanghai and Guangdong, where the machine learning models perform particularly well.

After comparing the MAE and RMSE of the carbon trading markets in each region, it is possible to identify the regions with the most serious problems in market operation mechanisms. Tianjin and Chongqing are particularly prominent, showing large MAE and RMSE values of 7.23 and 12.31, and 8.76 and 14.07, respectively.These high error metrics indicate that the carbon markets in these two regions suffer from significant uncertainty and volatility in price forecasting, suggesting that there are a number of imperfections in the operating mechanisms of the two markets that need to be further improved. .

Of the three machine learning models, the Gradient Booster performed the best. In several regions, the MAE and RMSE of the Gradient Booster are lower than the other models. For example, in Shenzhen, the MAE of the gradient boosting machine is 5.75 and the RMSE is 7.86, with an accuracy improvement of 49.89% and 46.01%, respectively. This trend is also evident in Shanghai and Guangdong, showing that the gradient lifter is more capable of capturing complex patterns in the data.

4.3. Eigenvalue significance

Among all machine learning models, Gradient Boosting Machine (GBM) shows the most excellent performance and becomes an effective tool for carbon pricing prediction. In order to delve into the important factors affecting carbon pricing and to further reduce pricing errors, it is necessary to demonstrate in detail the eigenvalues of the Gradient Boosting Machine for each region. This process will help to identify and analyze the degree of contribution of each feature in the model prediction, thus revealing the key drivers affecting the price volatility of carbon trading.

Table5 Ranking of importance of eigenvalues by region

Shenzhen			Shanghai			Beijing		
rankings	diagnostic property	significance	rankings	diagnostic property	significance	rankings	diagnostic property	significance
1	LNGP	0.1409	1	BY	0.2382	1	BY	0.1895
2	SP	0.1267	2	SP	0.1418	2	RCP	0.1067
3	COP	0.1051	3	EI	0.0892	3	IOP	0.0944
4	MI	0.0968	4	LNGP	0.0640	4	SP	0.0886

ISSN: 2710-0170

5	DJIA	0.0962	5	RCP	0.0608	5	SZCI	0.0816
6	BY	0.0875	6	COP	0.0598	6	LNGP	0.0720
7	IOP	0.0715	7	MI	0.0543	7	II	0.0703
8	EI	0.0649	8	DJIA	0.0538	8	NCI	0.0537
9	NCI	0.0479	9	SZCI	0.0521	9	SCI	0.0525
10	II	0.0467	10	NCI	0.0433	10	EI	0.0493
11	RCP	0.0455	11	II	0.0400	11	DJIA	0.0459
12	SZCI	0.0380	12	SCI	0.0396	12	MI	0.0389
13	SCI	0.0255	13	IOP	0.0334	13	GP	0.0284
14	GP	0.0069	14	GP	0.0296	14	COP	0.0281

Continued from table 5

Guangdong			Tianjin			Hubei			Chongqing		
rankings	diagnostic property	significance	rankings	diagnostic property	significance	rankings	diagnostic property	significance	rankings	diagnostic property	significance
1	BY	0.3378	1	COP	0.1335	1	BY	0.2022	1	COP	0.1104
2	COP	0.1205	2	II	0.1025	2	COP	0.1223	2	SCI	0.1102
3	LNGP	0.0968	3	EI	0.1010	3	EI	0.1007	3	BY	0.1000
4	SCI	0.0729	4	LNGP	0.0857	4	DJIA	0.0868	4	LNGP	0.0916
5	SZCI	0.0590	5	SZCI	0.0780	5	SCI	0.0689	5	EI	0.0838
6	NCI	0.0500	6	MI	0.0762	6	SZCI	0.0673	6	NCI	0.0827
7	DJIA	0.0489	7	IOP	0.0761	7	NCI	0.0575	7	GP	0.0730
8	II	0.0461	8	SP	0.0697	8	LNGP	0.0531	8	RCP	0.0659
9	EI	0.0429	9	BY	0.0558	9	II	0.0525	9	SP	0.0574
10	IOP	0.0418	10	RCP	0.0545	10	MI	0.0458	10	II	0.0541
11	RCP	0.0320	11	NCI	0.0475	11	GP	0.0437	11	MI	0.0526
12	GP	0.0191	12	GP	0.0432	12	IOP	0.0359	12	SZCI	0.0503
13	MI	0.0185	13	SCI	0.0421	13	SP	0.0347	13	IOP	0.0458
14	SP	0.0138	14	DJIA	0.0343	14	RCP	0.0284	14	DJIA	0.0223

After analyzing the eigenvalues of the gradient boosters in different regions, it can be observed that there are significant differences in the key factors affecting carbon pricing forecasts in each region, reflecting the uniqueness of each regional market and its sensitivity to carbon pricing. For example, carbon pricing in Shenzhen is mainly influenced by LNG prices and steel prices, showing its high sensitivity to energy price fluctuations, while in Shanghai, bond returns are the main influencing factor, demonstrating the significant role of financial markets in carbon pricing. Beijing and Guangdong are similarly influenced by bond returns and crude oil prices, reflecting the market dependence of traditional energy prices. Carbon pricing in Tianjin, on the other hand, is closely related to industrial indices and crude oil prices, suggesting that we

ISSN: 2710-0170

should pay attention to the impact of industrial activities on energy demand. Carbon pricing in Hubei and Chongqing is also influenced by both financial markets and energy prices, emphasizing the importance of market sentiment and energy price volatility. Therefore, future carbon pricing strategies should be tailored to the characteristics of different regions, focusing on energy prices, financial indicators and industrial activities, in order to improve the accuracy of forecasts, reduce pricing errors and promote the healthy development of the carbon market.

5. Robustness tests

5.1. Five-fold cross-validation

In this study, a training, validation, and testing approach to data partitioning was used, where the last year in the sample was used as the test set[10] . For each algorithm, this paper uses a five-fold cross-validation method to construct a machine learning model and uses the leave-out set to assess the actual effectiveness of the model. Five-fold cross-validation is a widely used resampling technique in machine learning to estimate the performance of a model on a limited sample of data.

Specifically, in this paper, the observations in the cross-validation sample are randomly divided into five equal-sized subsets. Each subset is sequentially used as the validation set, and the remaining four subsets are combined as the training set. The model is constructed based on the training set and evaluated on the validation set using a variety of evaluation metrics. A comprehensive assessment of the model performance is generated by averaging the results of the five validation sessions. Thus, all observations are used for training and validation, and each observation is used only once as validation data.

The model constructed from the cross-validation process was then applied to the test set following the training and validation periods to generate loss estimates for that period. The observations in the test set were not involved in the model development process. Therefore, the test set was used only for final evaluation and was not used as a criterion for model selection.

In this paper, the carbon trading prices in the three years of 2021, 2022 and 2023 are taken as the validation set, and the training set is divided into three time periods of 2014-2020, 2014-2021 and 2014-2022, respectively, to this predict the carbon trading prices in each year. By comparing the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) of the prediction results, if the prediction results are consistent with the above experimental results, the experimental results can be considered robust. This methodology is designed to assess the predictive ability of the model over different time periods to ensure its reliability and validity in practical applications.

5.2. Robustness test results

Table 6 Five times spread validation results

Areas	validation set	traditional linear regression		decision tree		random forest		Gradient lifts	
Shenzhen		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
	2021	10.09	13.62	14.12	16.35	10.68	11.48	8.06	8.97
	2022	33.33	38.20	21.49	27.11	21.49	25.16	20.43	25.22
	2023	58.38	61.76	19.68	21.53	16.46	16.87	16.39	17.63
Shanghai									
	2021	3.34	4.02	1.53	1.64	7.52	8.87	5.38	5.87
	2022	18.91	22.51	14.17	15.48	14.01	15.30	15.04	16.16
	2023	14.78	16.26	8.48	13.19	7.63	8.78	7.89	9.60
Beijing									
	2021	43.88	45.58	22.48	29.89	22.20	28.86	22.26	28.38
	2022	16.79	18.78	20.19	26.15	18.70	22.65	20.42	25.52
	2023	29.47	33.27	16.04	17.75	11.34	13.84	10.52	12.12

Guangdong									
	2021	12.04	15.85	10.51	11.79	11.96	13.15	11.18	12.41
	2022	34.40	35.16	27.52	29.06	33.37	33.78	29.84	30.24
	2023	30.09	31.95	20.09	24.55	17.34	21.81	14.47	21.07
Tianjin									
	2021	16.76	19.88	5.25	5.71	11.11	11.42	5.73	6.20
	2022	22.54	23.67	6.57	7.79	6.51	7.59	7.30	8.59
	2023	10.97	12.08	9.31	13.45	6.30	8.61	5.75	7.70
Hubei									
	2021	15.82	18.28	7.53	8.97	4.99	6.70	6.98	8.59
	2022	5.36	6.40	13.36	13.41	10.73	11.01	11.96	12.32
	2023	7.98	10.00	1.69	2.38	3.98	4.69	2.83	3.39
Chongqing									
	2021	10.49	13.14	8.63	9.78	11.37	12.19	11.19	11.99
	2022	10.53	11.64	7.16	8.31	6.53	7.80	6.14	6.93
	2023	5.40	7.06	3.49	4.03	3.00	3.64	6.96	7.82

As can be seen from the results of the five-fold cross-validation, the experimental results show good robustness. Among several regions, the random forest and gradient boosters show significant accuracy improvement compared to traditional linear regression, especially in Shenzhen and Shanghai, where the accuracy improvement reaches 48.87% and 63.37%, respectively. In addition, the model performances in Guangdong and Hubei regions also show high accuracy enhancement, especially in Random Forest and Gradient Booster, where both MAE and RMSE are significantly lower than those of traditional linear regression, reaching 74.37% and 74.73% enhancement, respectively. These results indicate that the selected models have strong adaptability and predictive ability when dealing with data from different regions, reflecting the robustness of the models as a whole.

6. Conclusion and suggestiongs

Based on the data of seven major carbon trading markets in China from 2014 to 2023, this paper analyzes the impact of macroeconomic factors on carbon trading prices, and adopts a machine learning method to provide an in-depth analysis of the pricing problem in the carbon trading market. The results show that there is a significant negative correlation between macroeconomic factors and carbon trading price, and the traditional linear regression model is affected by data frequency and other factors when predicting the carbon trading price, which leads to inaccurate prediction results. By introducing machine learning models such as decision trees, random forests and gradient boosters, this paper significantly improves the accuracy of carbon trading price prediction, especially in the characterization of the market in different regions, which reveals the key influencing factors of carbon pricing in each region.

The research in this paper provides certain reference significance for governing the imbalance of the carbon trading market system and optimizing the carbon pricing model. First, strengthen regional collaboration and policy coordination. When promoting low-carbon economic transformation, regions should fully recognize the spatial spillover effect of energy consumption on regional low-carbon transformation. In order to realize the low-carbon transformation of the regional economy, it is necessary to combine the characteristics of the region and formulate precise policies according to local conditions. At the same time, it should consider the areas of industrial chain synergy, industrial layout, energy production and consumption in an integrated manner, strengthen cooperation with neighboring regions, and establish inter-regional linkage mechanisms, so as to avoid the negative impacts of the spatial spillover effect of inter-regional energy consumption. Second, optimize the energy consumption structure. Each region should pay attention to the role of carbon trading pilot policies in promoting low-carbon transformation, reasonably control the total energy consumption in the region, and optimize the energy consumption structure. Through the scientific formulation of supporting policies, enterprises should be guided to strengthen the independent innovation of energy utilization technology, and gradually offset the increased cost of participating in the carbon market exchange, so as to realize the transformation of regional low-carbon economy. Third, enhance market transparency and supervision. In order to promote the healthy development of the carbon market, policymakers should strengthen the regulation of the carbon market, enhance market transparency and ensure information symmetry to reduce the possibility of market manipulation. The efficiency of the carbon trading market can be further improved by establishing a sound market mechanism and enhancing the trust of market participants. Fourth, utilize advanced technology to enhance forecasting capabilities. It is recommended that policymakers and market participants utilize advanced technologies, such as machine learning, to enhance the forecasting ability of carbon trading prices. This not only helps to better understand market dynamics, but also provides a scientific basis for policy adjustment and market decision-making.

Acknowledgments

This work is supported by Project Grant for Undergraduate Research and Innovation Fund of Anhui University of Finance and Economics (Grant No: XSKY24012ZD).

References

[1] Gao K, Zhao Yi, Hu Bin. Can carbon emissions trading improve regional environmental pollution? --A synthetic control method test based on seven pilot provinces and cities[J]. Operations Research and Management,2024,33(04):194-199.

[2] LUO Liangwen,SUN Lixue,WANG Chen. Impact of carbon emissions trading pilot on low-carbon transformation of urban industry[J]. East China Economic Management,2024,38(07):39-52.DOI:10.19629/j.cnki.34-1014/f.231218001.

[3] Sun Xia,Liang Hongzhi. "Carbon Trading":A New Channel for Local Revenue Enhancement--A Quasi-Natural Experiment Based on China's Carbon Emission Trading Pilot Policy[J]. Academic Exploration,2024,(11):132-141.

[4] WU Zhenni, YIN Yingkai, JIN Ming. Study on the implied inter-regional carbon emission transfer responsibility under the policy spillover effect of China's carbon trading pilot[J]. Statistics and Information Forum,2024,39(07):82-96.

[5] XU Junwei,LIU Zhihua. Carbon trading pilot policy, energy consumption and regional low-carbon economic transformation[J]. Statistics and Decision Making,2024,40(20):172-177.DOI:10. 13546 /j.cnki.tjyjc.2024.20.030.

[6] Zhou Liang. Interest rate structure, market frictions, and intertemporal arbitrage-a machine learning-based prediction[J]. Statistics and Decision Making,2022,38(22):142-147.DOI:10. 13546/j. cnki.tjyjc.2022.22.027.

[7] Chen,J.;Peng,D.;Liu,Z.;Wu,L.;Jiang,M.ASustainableModelforForecastingCarbonEmissionTradingPrices.Sustainability2024,16,8324.https:/ /doi.org/10.3390/su16198324

[8] Fang, C.; Wang, W.; Wang, W. TheImpactofCarbonTradingPolicyonBreakthroughLow-Carbon Technological Innovation. Sustainability 2023,15,8277. https://doi.org/10.3390/su15108277

[9] Fang, C.; Wang, W.; Wang, W. TheImpactofCarbonTradingPolicyonBreakthroughLow-Carbon Technological Innovation.Sustainability2023,15,8277. https://doi.org/10.3390/su15108277

[10] Kexing, D., Baruch, L., Xuan, P., Sun, T., & Vasarhelyi, M. A. (2020).Machine learning improve saccountingestimates. Evidence from insurance payments. ReviewofAccountingStudies, 25(3), 1098-1134. https://doi.org/10.1007/s11142-020-09546-9

ISSN: 2710-0170

[11] NDRC,China2050HighRenewableEnergyPenetrationScenarioandRoadmapStudy.Availableonline:https://www.efchina.org/Attachments/Report/ report-20150420(accessedon6October2022).

[12] Yin,Y.,Jiang,Z.,Liu,Y.,&Yu,Z.(2019).FactorsAffectingCarbonEmissionTradingPrice:EvidencefromChina. EmergingMarketsFinanceandTrade,55(15),3433-3451.https:// doi.org/ 10.1080/ 1540496X.2019.1663166

[13] Zeng,S.;Fu,Q.;Yang,D.;Tian,Y.;Yu,Y.TheInfluencingFactorsoftheCarbonTradingPrice:ACaseofChinaagainsta "DoubleCarbon "Background. Sustainability 2023,15,2203. https://doi.org /10.3390/ su15032203

[14] Zhao,Y.,Zhao,H.,Li,B.etal.Pointandintervalforecastingforcarbontradingprice:acaseof8carbontradingmarketsinChina. EnvironSciPollutRes30,49075-49096(2023).https://doi.org/10.1007/s11356-023-25151-0