Article

Carbon Price Combination Forecasting Model Based on Lasso Regression and Optimal Integration

Yumin Li 1,†, Ruiqi Yang 1,†, Xiaoman Wang 2, Jiaming Zhu 2,*, and Nan Song 2

1 SILC Business School, Shanghai University, Shanghai 201800, China; yuminli@shu.edu.cn (Y.L.); yangrq5808@163.com (R.Y.)
2 School of Internet, Anhui University, Hefei 230039, China; xiaomanw0@163.com (X.W.); 18196616151@163.com (N.S.)
* Correspondence: jiamingzhu@ahu.edu.cn; Tel.: +86-18355150818
† These authors contributed equally to this work.

Abstract: Accurate carbon price index prediction can delve deeply into the internal law of carbon price changes, provide helpful information to managers and decision makers, as well as improve the carbon market system. Nevertheless, existing methods for combination forecasting typically arbitrarily choose a certain set of single forecasting models. However, a particular selection of forecasting models do not apply to all data sets due to the nonlinearity and nonsmoothness of the carbon trading price series. Therefore, choosing suitable single forecasting models for the combination is crucial. Considering the limitations of the current study, this study constructs a combined carbon trading forecasting model based on Lasso regression and optimal integration. By invoking the Lasso regression model, we can select suitable single forecasting models for combination forecasting based on the variation patterns of different training sets. Meanwhile, ARIMA, NARNN, LSTM, and 11 other single forecasting models are screened in this study, including both traditional statistical forecasting models and artificial intelligence forecasting models. First, the carbon price index is predicted using 11 single prediction models. Furthermore, given the multi-collinearity of the single prediction series, this study employs Lasso regression to reduce the dimensions of the single prediction models, which are then used to construct an optimal combination prediction model. Finally, the proposed model is applied to SZA-2017 and SZA-2019 carbon price data in Shenzhen. The results demonstrate that the model developed in this study outperforms other benchmark prediction models in terms of prediction error and direction accuracy, showing the efficacy of the proposed method.

Keywords: carbon price forecasting; lasso regression; optimal combined forecasting model

1. Introduction

Extreme weather events including hurricanes, floods, and droughts have become more common in recent years, and the accompanying issues of rising sea levels, food security, and diminished biodiversity pose severe challenges to productive human existence [1]. Due to the severity of the climate, immediate and extensive action is needed to reduce the global warming process. After the Kyoto Protocol was signed at the United Nations Climate Change Conference (UNFCC) in 1997, the Paris Agreement adopted at the 2015 Paris Climate Change Conference broke the negotiation deadlock and reached an extremely significant milestone in terms of the general arrangements for the international community to address climate change after 2020 [2]. Since then, representatives from various nations have discussed issues such as achieving emission reduction goals and creating clean energy at the UN Climate Change Conference every year to further implement the Paris Agreement. At the same time, they will also look into more all-encompassing and coordinated ways to address climate change. In particular, economists believe that reasonable carbon pricing is the most effective means of achieving this goal [3]. A carbon trading system can guide companies to adopt more ecological behaviors, effectively reduce social costs of emission...
reductions, drive low-carbon technology innovation, as well as investment financing, and allow for better policy choices [4]. Since 2013, China has selected eight provinces and cities, including Beijing, Shanghai, Tianjin, Chongqing, Hubei, Guangdong, Fujian, and Shenzhen, as carbon trading pilots and opened a national carbon trading market in 2021.

In the carbon trading market, carbon prices are affected by various complex factors, such as changes in supply and demand, thus being nonlinear and unstable [5]. Therefore, the accurate prediction of carbon prices can dig deeper into the intrinsic fluctuation law of carbon prices, accelerating the establishment of an efficient and perfect price system [6]. As a result, establishing a high-precision carbon price prediction model is essential.

Nevertheless, existing methods for combination forecasting typically use one or a few specific methods. Specifically, various forecasting methods can provide practical information for carbon price forecasting from different perspectives due to the varying model settings. This study combines various models for carbon price forecasting. Eleven single forecasting models, namely, the autoregressive integrated moving average model (ARIMA), the exponential smoothing prediction method (ES1), Holt’s exponential smoothing prediction method (ES2), the generalized autoregressive conditional heteroscedasticity prediction model (GARCH), the generalized regression neural network prediction model (GRNN), the long short-term memory prediction model (LSTM), the nonlinear autoregression with external input prediction model (NARX), the nonlinear autoregressive neural network prediction model (NARNN), the least square support vector machine prediction model (LSSVM), the recurrent neural network model (RNN), and the back propagation neural network prediction model (BPNN), are employed to fully exploit the information provided by each method and establish the optimal combination forecasting model based on the least absolute shrinkage and selection operator (Lasso) regression.

The qualitative analysis method is commonly used for the first investigation of carbon price forecasting [7]. With the swift progress of statistics and economics, as well as the rapid penetration of machine learning technology, researchers are gradually exploring more accurate methods to capture the evolution of carbon price series. In general, the current research on carbon price forecasting is slowly shifting focus from single forecasting methods to hybrid forecasting methods.

Specifically, the single prediction method mainly refers to classical measurement and excellent artificial intelligence models. The former is based on a traditional econometric model to predict the time series. For instance, ARIMA, the most traditional time series prediction model, was used to forecast the linear components of the European Union’s carbon pricing series with favorable performance [8]. In addition, based on the second-stage carbon price data of the European Climate Exchange (ECX) from 2008 to 2011, GARCH [9,10] was found to have a better fitting degree than the k-nearest neighbor model and implied volatility. Furthermore, the improved ARIMA-GARCH model also made a good prediction of the carbon price fluctuation in the European Union, which effectively captured the skewness and the fluctuation behavior in different stages [11]. To sum up, these models are all based on comprehensive statistics and can make sound predictions for stationary time series under certain assumptions. Notwithstanding, considering the prominent non-stationary and nonlinear features of carbon trading price series, the econometric prediction model cannot capture such elements accurately, so the prediction accuracy needs further improvement. Fortunately, the artificial intelligence model makes up for this defect to some extent.

With the swift advancement of machine learning, a growing number of artificial intelligence models are being used to forecast carbon prices. Compared with the econometric prediction model, these models have more robust adaptability and generalization ability. Furthermore, they can more accurately mine the nonlinear laws between data. The BP neural network, for example, is an artificial neural model that conducts reverse training based on errors, can modify any complexity mapping relationship, and has a strong nonlinear prediction ability [12]. Nevertheless, it is easy for the model to suffer from premature convergence caused by local minima. To solve this problem, Sun and Huang [13]
combined the genetic algorithm (GA) with the BP model to further improve the training speed. At the same time, they applied this model to predict carbon price series in three Chinese markets, Beijing, Shanghai, and Hubei. Moreover, the prediction effectiveness of the model was proven using three error indicators. Likewise, LSSVM has also achieved good results in carbon price prediction, the core problem of which is determining the kernel function and the optimal parameters. In comparison to the BP neural network, this model can successfully solve the volatility of training effectiveness. Zhu et al. [14] used empirical mode decomposition (EMD) to improve the LSSVM model’s resilience. In addition, Atsalakis et al. [15] creatively proposed three-carbon price prediction methods: the neuro-fuzzy hybrid controller (NFHC), the adaptive neuro-fuzzy inference system (ANFIS), and the artificial neural system. According to the findings, the second method has the highest forecast accuracy and the largest development potential. However, these artificial intelligence models also have their limitations, for example, that they cannot completely capture various feature information of time series. When the time series fluctuates wildly, this kind of model cannot deliver ideal prediction results.

On the other hand, considering that the single forecasting model has certain limitations and cannot accurately capture the practical characteristics of the carbon price series, the existing carbon price research gradually adopts the combination forecasting method [16]. This method may fully exploit the benefits of each individual prediction model, hence increasing forecast accuracy in particular. Zhu et al. [17] combined LSSVM, GARCH, and EMD to set up a new nonlinear integrated combination forecasting model. Firstly, the carbon price series was decomposed using EMD into a residual value and multiple intrinsic modal functions, which are classified as high-frequency, low-frequency, and trend functions. After this, they considered using the GARCH model to predict the high-frequency function and using the LSSVM model to predict the low-frequency as well as trend functions. Finally, they employed LSSVM to integrate the predicted values. Nevertheless, modal mixing may interfere with EMD decomposition, so functions with the same frequency cannot be accurately decomposed. For this reason, researchers created the variational modal decomposition (VMD) model [18]. By combining this model, modal reconstruction (MR), and the optimal combination forecasting model (CFM), Zhu et al. [19] developed a novel combination forecasting method that increased the carbon price prediction accuracy substantially. In addition, Sun et al. [20] also created a new integrated forecasting model, which combined weighted least squares support vector machine (WLS-SVM), linear particle swarm optimization (LDWPSO), and integrated empirical mode decomposition (EEMD). Moreover, they apply it to the empirical research of carbon trading markets in Hubei, Guangdong, and Shanghai. Similarly, by combining fractal Brownian motion (FBM) with GARCH, Liu et al. [21] successfully predicted the carbon trading price of the European Energy Exchange (EUA).

After carefully analyzing the current carbon price combination forecasting models, we can see that each individual prediction method can extract sufficient information from various perspectives. Nonetheless, the accuracy of each method’s forecast varies over time. That is, the “high and low” condition emerges. Consequently, combined forecasting can incorporate the benefits of each forecasting method while also improving model forecasting accuracy. However, because there are a host of single prediction methods, some of them supply repeated information, increasing prediction error and computational complexity if they are all utilized together. As a result, one of the significant difficulties in this study is effectively picking suitable approaches among the single prediction models for combination.

Generally, linear and nonlinear algorithms are the most common data dimensionality reduction techniques. Researchers studied linear dimension reduction methods such as principal component analysis (PCA), independent component analysis (ICA), and factor analysis (FA) in the early research [22–24]. These methods, based on feature extraction, translate data from a high-dimensional to a low-dimensional space and use the extracted variables to represent the original whole. Furthermore, scholars are gradually adopting
nonlinear dimensionality reduction approaches such as isometric mapping (Isomap), kernel-based principle component analysis (KPCA), and stochastic neighborhood embedding (SNE) [25–27] as data structures become more complicated. KPCA is a PCA-based method for improving nonlinear data processing [28]. Simultaneously, SNE is a fantastic learning tool with excellent visualization functions. Laurens et al. [29] devised the t-SNE model on this foundation. The algorithm pays more attention to the correlation between neighboring sample points to better demonstrate the nonlinear relationship between variables. While these solutions can be used for a wide range of problems, they do have some limitations. Other issues, such as the selection of sample size, inability to detect outliers, and estimation of intrinsic dimensionality, have not been addressed. Considering the nonlinearity of carbon pricing data and the possibility of multiple co-linearities in a single forecasting model, we chose to eliminate collinearity statistically. In particular, the Lasso regression model is used to reduce the data dimensionality.

The Lasso regression model, on the other hand, has become a popular linear regression model in recent years [30], which is widely used in computer science, biomedicine, engineering, and other fields [31–34]. The essence of this model is a kind of compressed estimation. By constructing a penalty function, the sum of absolute values of regression coefficients is less than a specific constant. Meanwhile, the sum of residual squares is minimized so that some coefficients are selected as 0. Finally, the coefficients of each explanatory variable are determined. Furthermore, scholars have gradually introduced the Lasso regression model into economics research. For example, Anuja et al. [35] explored which factor is the key to becoming an influencer on social platforms with this method. Theodore et al. [36] used the Lasso regression model to analyze bitcoin income. They concluded that both the gold return rate and search intensity have the most significant impact on the bitcoin return rate. Similarly, Liu et al. [37] used the Lasso method to determine the variables affecting the efficiency of green innovation in high-tech industries.

Furthermore, Lasso regression models have also begun to be used by current academics to establish the weights of the combined models, thus improving the prediction accuracy. Among these, Diebold et al. [38] propose an approach based on “partial equality LASSO” (peLASSO), which outperforms other models in terms of forecasting results by decreasing the weights of the remaining portfolio weights to the average level and setting some weights to zero. This study was even applied to the European Central Bank Survey of Professional Forecasters, demonstrating its potential for practical application. In addition, Zhang et al. [39] compared the Lasso model with other models, such as ridge regression, principal component regression, and partial least squares in forecasting oil prices. The experimental findings demonstrate the Lasso model’s capacity to choose reliable predictors that add information beyond that offered by rival models. Similar to this, Zhang et al. [40] suggested using a time autocorrelation-based regression model to estimate the concentration of suspended sediment. By using the “shrinkage” effect of the Lasso model, the model was able to identify the most significant factor among the analyzed parameters and compress the contributions of other factors to zero. The results show that the Lasso method is able to obtain prediction results with minimum root mean square error and standard deviation. In conclusion, these studies demonstrate that the use of the Lasso model can increase forecasting accuracy and identify the most crucial predictors, giving decision makers more precise and dependable forecasting findings. In particular, no one has used this method to forecast carbon prices.

Accordingly, this research presents a Lasso regression-based carbon price combination forecasting model. First, nine well-behaved single prediction models are selected to separately predict the carbon price series. Thus, the initial prediction values are obtained. In addition, the optimal weights of each model can be derived after filtering each model using a Lasso regression-based combined forecasting model. Lastly, the product of the predicted values of each model and the optimal weight is added to obtain the final combined prediction result. In addition, we use the carbon price data of Shenzhen for empirical research and consider four error indicators to corroborate the model’s prediction effectiveness.
This study contributes to the literature in following ways:

(1) For the first time, the Lasso regression model is used to investigate carbon price combination prediction, which partially solved the challenge of calculating the weight of the combination forecast.

(2) Past research on carbon price combination forecasting usually considered three or four single forecasting models. In comparison, eleven excellent single prediction models are selected in this study, which can capitalize on the advantages of different carbon price prediction models and internalize the characteristics of carbon price series more comprehensively, thus improving the prediction accuracy.

The remainder of this research is as follows. Section 2 introduces eleven commonly used single prediction models and Lasso regression models, which lays a solid theoretical foundation for proposing new models later. Then, the combination forecasting model of the carbon price based on Lasso regression is constructed. Additionally, Section 3 describes the data and the evaluation index. Section 4 contains the empirical analysis. Finally, Section 5 elaborates on the study’s primary findings and discusses future research.

2. Methodology

Several excellent single prediction models and Lasso regression models are introduced in this part, laying a solid theoretical foundation for the new model. On the basis of these former models, we establish a carbon price combination forecasting model based on Lasso regression.

2.1. A Meta-Analysis of Carbon Price Portfolio Forecasting Models

This article conducts a meta-analysis of the literature on carbon price forecasting for the years 2000–2023. First, we conducted searches for the terms “carbon price forecasting” and “combination forecasting model” in databases including Google Scholar, Web of Science, and Science Direct. As a result, we found a total of 606 papers. Second, the following criteria were established to further screen the literature, taking into account the needs for a meta-analysis for literature screening and the alignment of the topic with the research direction: excluding literature whose subject matter did not correspond to the research question; keeping only one piece of literature that had a great deal of overlap in the research models and conclusions; and excluding empirical studies, theoretical studies, and review studies that lacked conversion indicators for effect values. Finally, we selected 56 papers for independent coding and conversion into meta-analysis effect values. Specifically, the coding content consisted of the description of the study and the effective value, in order. The former included the name, number, author, year, and journal title of the study, while the latter included the study model, sample size and its related properties, and other effect values. The entire process is shown in Figure 1.

The meta-analysis’s conclusions were reached after a number of analyses, including heterogeneity tests and bias analyses. Initially, we discovered that, among the screened studies on carbon trading price forecasting, the prediction outcomes of the literature using combination forecasting models were significantly superior to those of the literature using single forecasting models. Additionally, for various datasets, each model’s accuracy and prediction outcomes vary, each having benefits of their own. Consequently, it is crucial to choose an appropriate combination of single carbon price forecasting models. At last, we chose 11 single forecasting models from the literature that are relatively well-fitted and appear most frequently in high-quality journals as the basis for the following combination of projections. These models were chosen from a variety of perspectives.
Through the databases of Google Scholar, Web of Science, and Science Direct, a total of 606 papers were collected.

517 duplicates were removed.

89 articles were included in the literature.

After reading the titles and abstracts, 28 papers were discarded.

Read and assess all 61 documents in every detail.

Five studies were disregarded, primarily due to their low caliber and dearth of results.

In conclusion, 56 pieces of literature were incorporated into the research.

**Figure 1.** Flow chart of the meta-analysis.

### 2.2. Several Well-Behaved Single Prediction Models

#### 2.2.1. Exponential Smoothing Prediction Method

This method reduces the influence of distant data on the prediction results by assigning the weight used to converge [41], thus enhancing the prediction accuracy. The first exponential smoothing prediction approach can be utilized with no evident change rule but some correlation in the time series. The basic formula is as follows:

\[ S_t^{(1)} = a \cdot y_t + (1 - a) \cdot S_{t-1}^{(1)} \]  

In this study, \( S_t^{(1)} \) is expressed as the predicted carbon price, \( y_t \) denotes the actual carbon price data in the current period, and \( S_{t-1}^{(1)} \) represents past period exponential smoothing values. The smoothing coefficient, \( a \), has a value range of \([0, 1]\).

#### 2.2.2. Holt’s Exponential Smoothing Prediction Method

The first exponential smoothing prediction will lag when the time series shows a large linear trend. As a result, Holt’s exponential smoothing prediction is used to correct the variance based on the first exponential smoothing. The formula is as follows:

\[ S_t^{(2)} = aS_t^{(1)} + (1 - a)S_{t-1}^{(2)} \]

\[ y_{t+T} = a_t + b_t \cdot T, \]

\[ \begin{align*}
  a_t &= 2S_t^{(1)} + S_{t-1}^{(2)} \\
  b_t &= \frac{a}{1-a} \left( S_t^{(1)} - S_t^{(2)} \right)
\end{align*} \]

where \( S_t^{(1)} \) and \( S_t^{(2)} \) respectively indicate the first and second exponential smoothing prediction values of the \( t \)-period carbon price series and \( S_{t-1}^{(2)} \) presents the \( t-1 \) period carbon pricing series’ second exponential smoothing prediction result. In addition, \( y_{t+T} \) denotes the \( t+T \) period prediction value, \( T \) means the number of periods from the first to the second prediction period, and \( a_t \) and \( b_t \) are intercept and slope, respectively.
2.2.3. Generalized Autoregressive Conditional Heteroscedasticity Prediction Model (GARCH)

GARCH is an extended form of the ARCH model that is widely used to predict and analyze the volatility of financial time series [42]. The following is the specific calculating formula:

\[ y_t = x_t \theta + \epsilon_t, \epsilon_t \sim N(0, \sigma_t^2) \]  \hspace{1cm} (4)

\[ \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2 \]  \hspace{1cm} (5)

Formula (4) is the conditional mean equation, where \( y_t \) denotes the predicted value of carbon price, \( x_t \) means the independent variable, and \( \epsilon_t \) is the random disturbance term. Meanwhile, Formula (5) shows the conditional variance formula, where \( p \) is lag order and \( q \) denotes the residual square and lag term number. In addition, both \( \alpha_i \) and \( \beta_j \) represent regression coefficients, and \( u_i \) is a sequence of independent identically distributed random variables with mean 0 and variance 1.

To meet the positive variance and ensure the stability of the model, the following conditions should be met: \( \alpha_0 > 0, \alpha_i \geq 0, \beta_i \geq 0, \sigma_0^2 + \sum_{j=1}^{q} \beta_j > 0 \), and \( \alpha_0 + \sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1 \).

2.2.4. Autoregressive Integrated Moving Average Prediction Model (ARIMA)

ARIMA is a classical time series model proposed by Box–Jenkins [43], widely used to predict various time series. Assume that the variance of the disturbance terms is constant and the residual series are independent of each other. The formula is expressed as follows:

\[ \omega_t = c + \Phi_1 \omega_{t-1} + \cdots + \Phi_p \omega_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} \]  \hspace{1cm} (6)

\[ \phi(L)(1 - L)^d y_t = c + \Theta(L) \epsilon_t \]  \hspace{1cm} (7)

\[ \left\{ \begin{array}{l}
\Phi(L) = 1 - \Phi_1 L - \Phi_2 L^2 - \cdots - \Phi_p L^p \\
\Theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \cdots - \theta_q L^q \\
\omega_t = \Delta^d y_t = (1 - L)^d y_t
\end{array} \right. \]  \hspace{1cm} (8)

In the above formula, \( y_t \) represents the carbon price counted at the time \( t \), and \( \omega_t \) indicates the stationary sequence of \( y_t \) after \( d \) times of difference. In addition, \( \epsilon_t \) is the variance, \( L \) is the delayed operator, and \( \theta \) is the moving average model’s order. Moreover, \( \Phi_1, \Phi_2 \cdots \Phi_p \), as well as \( \theta_1, \theta_2 \cdots \theta_q \), are unknown parameters to be estimated in the model.

2.2.5. Least Square Support Vector Machine Prediction Model (LSSVM)

LSSVM is an optimized support vector machine (SVM). In this model, equality constraints are used instead of inequality constraints, thus transforming the original problem into solving linear equations, which dramatically improves the learning efficiency [44]. The optimization function is as follows:

\[ \left\{ \begin{array}{l}
\text{min}_{\omega, \mu, \epsilon} \frac{1}{2} \|\omega\|^2 + \frac{1}{2} \lambda \sum_{i=1}^{N} \epsilon_i^2 \\
\text{s.t.} \quad y_i = \omega^T \psi(x_i) + m + \epsilon_i, i = 1, 2, \cdots, N
\end{array} \right. \]  \hspace{1cm} (9)

where \( \lambda \) presents the regularization parameter, which directly determines the model’s training error and goodness of fit, and \( \epsilon_i \) is the error variable. Next, the Lagrange multiplier \( a_i \) is introduced to convert the aforementioned function into the Lagrange form:

\[ L = \frac{1}{2} \|\omega\|^2 + \lambda \sum_{i=1}^{N} \epsilon_i^2 - \sum_{i=1}^{N} a_i \left( \omega^T \psi(x) + m + \epsilon_i - y_i \right) \]  \hspace{1cm} (10)
Find the optimal solution of the above formula, and perform partial differentiation on \( \omega, m, e_i, a_i \). After that, eliminating \( e_i \) and \( \omega \) yields the following linear equations:

\[
\begin{pmatrix}
0 \\
R_N \Omega_{ij} + \frac{1}{2} I
\end{pmatrix}
\begin{pmatrix}
m \\
a
\end{pmatrix}
= \begin{pmatrix}
0 \\
y
\end{pmatrix}
\] (11)

where \( \Omega_{ij} = \psi(x_i)^T \psi(x_i) = h(x_i, x_j) \) for \( i, j = 1, 2, \ldots, N \)
\( R_N = [1, 1, \ldots, 1]^T, a = [a_1, a_2, \ldots, a_N]^T \).

The values of \( a \) and \( m \) can be obtained by solving the above equation, from which the prediction function of LSSVM is derived:

\[
\begin{cases}
y = \sum_{i=1}^N a_i H(x, x_i) + m \\
H(x, x_i) = \exp\left(-\frac{(x_i - x)^2}{2\sigma^2}\right)
\end{cases}
\] (12)

2.2.6. Generalized Regression Neural Network Prediction Model (GRNN)

GRNN belongs to the forward propagation network with excellent learning efficiency and nonlinear mapping ability [45]. As a result, it is commonly employed in predicting nonlinear time series. In addition, this model has four layers: input, mode, summation, and output. The image is shown in Figure 2.

Figure 2. The flowchart of GRNN.

The formula of each layer is as follows:

\[
\begin{cases}
m_i = \exp\left[-\frac{(X - X_i)^T (X - X_i)}{2\xi^2}\right], i = 1, 2, 3, \ldots n \\
\sum_{i=1}^n y_i \exp\left[-\frac{(X - X_i)^T (X - X_i)}{2\xi^2}\right]; S_{Nj} = y_i M_i, j = 1, 2, \ldots, k
\end{cases}
\] (13)

\[
\sum_{i=1}^n \exp\left[-\frac{(X - X_i)^T (X - X_i)}{2\xi^2}\right]; S_D = \sum_{i=1}^n M_i
\] (14)

\[
y_j = \frac{S_{Nj}}{S_D}, j = 1, 2, \ldots, k
\] (15)
where \( \sigma \) presents the smoothness factor and \( X_i \) denotes the learning sample of the \( i \)th neuron. Additionally, the connection weight between the \( i \)th model layer neuron and the \( j \)th summation layer neuron is \( y_{ij} \). Moreover, \( y_j \) indicates the \( j \)th neuron’s output values.

2.2.7. Nonlinear Autoregressive Neural Network Prediction Model (NARNN)

NARNN can be regarded as the nonlinearization of the classical time series prediction model. This model takes effects through the values of the first \( d \) times of the \( t \) time of the historical series [46]. In addition, the network structure includes the input layer, hidden layer, and output layer. The following is the calculation formula:

\[
z(t) = f(z(t-1), z(t-2), \cdots, z(t-d)) \tag{16}
\]

where \( [z(t-1), z(t-2), \cdots, z(t-d)]^T \) denotes the input variable and \( z(t) \) implies the output variable, i.e., the forecast value of the carbon price at \( t \). Meanwhile, \( d \) presents the order of the delay signal.

2.2.8. Nonlinear Autoregressive with External Input Prediction Model (NARX)

NARX is an externally input nonlinear autoregressive neural network. Specifically, as a nonlinear system identification tool, NARX can be utilized for nonlinear time series forecasting and real-world problems such as air-gas compression systems and predicting the hysteresis effects of bolted joint structures [47,48]. The image is shown in Figure 3.

![Figure 3. The flowchart of NARX.](image)

The model is described as follows:

\[
z(n+1) = f \left( z(n), \cdots, z(n-d_z+1); v(n-k), v(n-k+1), \cdots, v(n-d_v-k+1) \right) \tag{17}
\]

where \( v(n) \in \mathbb{R}, w(n) \in \mathbb{R}, z(n) \in \mathbb{R} \) represent the input and output in this model at \( n \) time points, respectively. In addition, \( d_v \geq 1, d_z \geq 1, d_v \leq d_z, d_v \) denotes the input-storage order, and \( d_z \) indicates the output-storage order. In addition, \( k(k \geq 0) \) is a lag term called process time.
2.2.9. Long Short-Term Memory Prediction Model (LSTM)

The LSTM neural network has been widely utilized in dealing with timing problems such as speech recognition and translation since it was established by Hochreiter and Schmidhuber in 1997 [49]. After submitting a new state for linear feedback transfer, the gradient explosion and gradient vanishing problems of recurrent neural networks can be effectively solved. The image is shown in Figure 4.

![Figure 4. The flowchart of LSTM.](image)

The expressions of the model are as follows:

\[
\begin{align*}
    z_t &= \sigma(W_f \cdot [m_{t-1}, x_t] + b_f) \\
    i_t &= \sigma(W_i \cdot [m_{t-1}, x_t] + b_c) \\
    \tilde{c}_t &= \tanh(W(c)x_t + U(e)m_{t-1} - 1) \\
    c_t &= z_t \ast c_{t-1} + i_t \ast \tilde{c}_t \\
    o_t &= \sigma(W_0 \cdot [m_{t-1}, x_t] + b_0) \\
    m_t &= o_t \ast \tanh(c_t)
\end{align*}
\]  

(18)

where \(x_t\) denotes the current input, \(m_{t-1}\) is the hidden state at the last moment, and \(x_t\) and \(m_{t-1}\) jointly determine the amount of information retained by the LSTM cell state. In addition, \(\sigma\) presents the logistic function, \(z_t\) indicates the forgetting gate, and \(o_t\) is the output gate. Moreover, \(W_f\) and \(b_f\) are the trainable parameters in the forget gate, \(\tilde{c}_t\) is the information added to the LSTM cell state, and \(c_t\) represents the updated LSTM cell state.

2.2.10. Back Propagation Neural Network Prediction Model (BPNN)

The BP neural network model shows good predictive ability in describing nonlinear problems, such as carbon price prediction [49]. Precisely, the basic structure is made up of three layers: input, concealed, and output. The model is described as follows: The output of each unit in the hidden layer and the output layer are:

\[
n_i = f(H_i), i = 1, 2, \cdots, n
\]

(19)

\[
y_i = f(O_i), i = 1, 2, \cdots, l
\]

(20)

Calculate the sum of squared errors according to the predicted value and expected output value \(d_i\):

\[
E = \frac{1}{2} \sum_i e_i^2 = \frac{1}{2} \sum_{i=1}^{l} (d_i - y_i)^2
\]

(21)
Adjust the connection weight by error:

\[ t_{ij}(t + 1) = t_{ij}(t) + \eta (dt - yt)yt(1 - yt)net_j, i = 1, 2, \cdots l, j = 1, 2, \cdots n \]  
(22)

\[ w_{ij}(t + 1) = w_{ij}(t) + \eta \sum_{k=1} \left( d_k - y_k \right)y_k(1 - y_k)t_{ki}net_j(1 - net_i)x_j 
\]
(23)

where the implicit and output layer input variables are denoted by \( H_i \) and \( O_i \). The weight between the output layer and the hidden layer is \( w_{ij} \). In addition, the weight between the hidden layer and the output layer equals \( t_{ij} \), and the function between the output layer unit and the hidden layer unit is a sigmoid function.

2.2.11. Recurrent Neural Network Model (RNN)

The recurrent neural network model (RNN) is a feedforward neural network model that is an upgraded model based on the BP neural network [50]. It effectively solves sequence data learning issues and is frequently used in speech recognition, machine translation, and other domains.

In contrast to the BP neural network model, this model forms connections between itself and neurons, allowing the network to have a memory function. The image is shown in Figure 5.

Figure 5. The flowchart of RNN.

The formula is as follows:

\[ R = \frac{1}{2} \sum_{i} (H_i - O_i)^2 \]  
(24)

\[ \Delta W_{kj}^h(n + 1) = -\alpha \frac{\partial E}{\partial W_{kj}} + \beta \Delta W_{kj}^h(n) \quad (k = 1,2,\cdots,K; j = 1,2,\cdots,J) \]  
(25)

where \( R \) denotes the mean square deviation between the actual and expected outputs \( O_i \) and \( H_i \). What is more, \( \partial \) and \( \beta \) represent learning efficiency and momentum factor, respectively. Simultaneously, \( n \) denotes the number of iterations, \( k \) is the number of inputs, and \( j \) indicates the number of neurons in the hidden layer.

2.3. Standard Lasso Regression Model

In general, the least squares regression approach (OLS) can be used to obtain the coefficients of the linear regression model by reducing the sum of squares of errors. However, when the quantity of explanatory variables is too large or partially linearly related,
the prediction accuracy of this model is often insufficient and the influence of collinearity cannot be eliminated, so we consider building a Lasso regression model.

The Lasso regression model can not only deal with the issue of collinearity but can also rationalize the coefficients of each explanatory variable more accurately. Its essence is a kind of compressed estimation. By constituting a penalty function, the sum of absolute values of regression coefficients is less than a constant. Moreover, the sum of residual squares is also minimized, so some coefficients are filtered as 0. The specific penalty function is as follows:

$$\hat{\beta}_{\text{Lasso}} = \arg \min_{\beta \in \mathbb{R}^d} \left( \| Y - X\beta \|^2 + \lambda \sum_{j=1}^{d} |\beta_j| \right)$$

(26)

The above formula can be rewritten as:

$$\hat{\beta}_{\text{Lasso}} = \arg \min_{\beta \in \mathbb{R}^d} \| Y - X\beta \|^2 \text{ s.t. } \sum_{j=1}^{d} |\beta_j| \leq t, t \geq 0$$

(27)

where the second formula is the Lagrange form of the first formula, and $t$ denotes the harmonic parameter. If $t_0 = \sum_{j=1}^{d} |\hat{\beta}_j^{(\text{OLS})}|$, when $t < t_0$, some coefficients of the penalty function will be 0, thus enhancing the model’s prediction performance. On the other hand, as a regularization parameter, $\lambda$ presents the core of the fitting effect. If its value is too large, most of the coefficients may be zero, thus decreasing the forecasting precision of the model. Otherwise, it will lead to over-fitting of the model. In this study, we consider the cross-validation method to select $\lambda$.

2.4. Combination Forecasting Model of Carbon Price Based on Lasso Regression

A combined forecasting model denotes that calculating different weights according to the forecasting values of individual models can obtain the eventual forecasting value [51]. During the whole process, the problem of gaining weight is crucial, which directly determines the prediction effect of the model. Therefore, how to obtain the optimal weight has become an important issue.

This study puts forward a combined forecasting model with the Lasso regression model that solves the dilemma of determining the optimal weight. At the same time, it makes up for the deficiencies in the existing carbon price prediction research. The complicated construction steps are shown under Figure 6:

![Figure 6. The framework of the proposed combination forecasting framework.](image-url)
Step 1: Data preprocessing.
First, fill in the blanks and delete outliers. Subsequently, test the nonlinearity and nonstationarity of the carbon price time series.

Step 2: Calculate the values of eleven single forecasting models.
According to the characteristics of different models, the optimal parameter values can be obtained in advance. After that, to obtain the initial forecast value, eleven single forecast models, such as ARIMA, NARNN, BP, and LSTM, are considered to predict the carbon price series. The prediction result is \( \hat{x}_{it}, i = 1, 2, \cdots, 10; t = 1, 2, \cdots, T \), which can be understood as the prediction value of the \( i \)th method at the \( t \) time.

Step 3: Correlation test and Lasso regression screening.
To test the correlation, the correlation coefficient between the expected value of the single forecasting model and the actual value of the carbon price is calculated. The Lasso regression model then selects a portion of the model coefficients as 0 based on the prediction outcomes of the single prediction model. These models will be phased out in the future. Furthermore, the chosen model will be used for the combination prediction to follow.

Step 4: The chosen model is employed for combination prediction.
The weighted average method is an effective way to increase combined forecasting accuracy. This strategy usually gives a constant weight to a single prediction model, but it ignores the fact that the forecast accuracy of each model varies with time. As a result, the induced ordered weighted averaging (IOWA) operator is used in this study, and combined prediction is gained by reducing the sum of error squares. Then, the predictions of various models are reordered at the same time. The modified sequence can then be weight-averaged to achieve the final anticipated value.

Assume that the real value of the carbon price at time \( t \) is \( x_t \). At the same time, the \( i \)th single prediction model’s prediction value \( x_{it} \) has a prediction accuracy of \( b_{it} \):

\[
b_{it} = \begin{cases} 0, & |(x_t - x_{it}) / x_t| \geq 1 \\ 1 - |(x_t - x_{it}) / x_t|, & |(x_t - x_{it}) / x_t| < 1 \end{cases}
\]

\[
\min J = \sum_{t=1}^{T} \mu_t^2 = \sum_{i=1}^{m} \omega_i (\sum_{t=1}^{T} \mu_{it} \mu_{jt})
\]

s.t. \[
\sum_{t=1}^{m} \omega_i = 1 \\
\omega_i \geq 0, i = 1, 2, \cdots, m
\]

\[
\mu_t = x_t - \hat{x}_t = \sum_{i=1}^{m} \omega_i (x_t - x_{b-index(it)}) = \sum_{i=1}^{m} \omega_i \mu_{it}
\]

where \( \mu_t \) represents the combined prediction error based on the IOWA operator, \( \mu_{it} \) denotes the single prediction method error, and \( \omega_i \) indicates the weight. In addition, the reordered sequence is represented by \( x_{b-index(it)} \). Consequently, we can obtain the ideal weight using this model.

Step 5: Compute the overall predicted value:

\[
\hat{x}'_t = \sum_{i=1}^{m} \omega_i' x_{b-index(it)}
\]

3. Data and Evaluation Indicators

3.1. Data Description

Since Shenzhen is one of the earliest pilot cities for carbon emissions trading in China, it has a better market system compared to other cities. Therefore, its carbon emission trading data are more comprehensive and representative. In this study, the Shenzhen SZA-2017 and SZA-2019 carbon price data sets from the CSMAR database are selected as experimental samples. The data range of SZA-2017 is from 31 July 2017 to 9 March 2022, with 427 data samples. The first 397 data sets are used as training sets, while the last 30 samples are used
as test sets. The data range of SZA-2019 is from 29 October 2019 to 7 March 2022, with
315 data points. Similarly, the test set consists of the latest 30 samples, while the training set
consists of the remaining data. Specifically, training samples are mostly used to determine
model parameters and regression coefficients, whereas test samples are primarily used to
assess the model’s prediction accuracy. The data set is shown in Table 1.

Table 1. Description of the datasets.

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Type</th>
<th>Size</th>
<th>Date Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>SZA-2017</td>
<td>Sample set</td>
<td>427</td>
<td>31 July 2017–9 March 2022</td>
</tr>
<tr>
<td></td>
<td>Training set</td>
<td>397</td>
<td>31 July 2017–28 July 2021</td>
</tr>
<tr>
<td></td>
<td>Test set</td>
<td>30</td>
<td>29 July 2021–9 March 2022</td>
</tr>
<tr>
<td>SZA-2019</td>
<td>Sample set</td>
<td>315</td>
<td>29 October 2019–7 March 2022</td>
</tr>
<tr>
<td></td>
<td>Training set</td>
<td>285</td>
<td>29 October 2019–13 December 2021</td>
</tr>
<tr>
<td></td>
<td>Test set</td>
<td>30</td>
<td>14 December 2021–7 March 2022</td>
</tr>
</tbody>
</table>

3.2. Forecasting Evaluation

We choose eight metrics to measure the model’s forecasting performance, which are
root mean square error (RMSE ), mean absolute percentage error (MAPE), sum of squares
error (SSE), mean fundamental error (MAE), Theil U statistic 1 (U1), index of agreement (IA),
success ratio (SR), and accuracy. The model’s prediction error is assessed using the first
four indices. The stronger the model’s prediction effect, the smaller the values of these
four indices are. Similarly, U1 is used to represent the model’s ability to make predictions.
The higher the prediction ability, the closer the index value is to 0. On the other hand, IA
SR represent the model’s generalization ability, and they are positively correlated with
the values. Additionally, the closer the accuracy is to 1, the better the model’s prediction
accuracy. The following is the calculating formula:

\[ \text{RMSE} = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (x_t - \hat{x}_t)^2} \]  
(32)

\[ \text{SSE} = \sum_{t=1}^{T} (x_t - \hat{x}_t)^2 \]  
(33)

\[ \text{MAE} = \frac{1}{T} \sum_{t=1}^{T} |x_t - \hat{x}_t| \]  
(34)

\[ \text{MAPE} = \frac{1}{T} \sum_{t=1}^{T} \frac{|x_t - \hat{x}_t|}{x_t} \]  
(35)

\[ U_1 = \frac{\sqrt{\frac{1}{T} \sum_{t=1}^{T} (x_t - \hat{x}_t)^2}}{\sqrt{\frac{1}{T} \sum_{t=1}^{T} x_t^2} + \sqrt{\frac{1}{T} \sum_{t=1}^{T} \hat{x}_t^2}} \]  
(36)

\[ \text{IA} = 1 - \frac{\sqrt{\frac{1}{T} \sum_{t=1}^{T} (\hat{x}_t - \bar{x}_t)^2}}{\sqrt{\frac{1}{T} \sum_{t=1}^{T} (\hat{x}_t - \bar{x}_t)^2} + \sqrt{\frac{1}{T} \sum_{t=1}^{T} (x_t - \bar{x}_t)^2}} \]  
(37)

\[ \text{Accuracy} = 1 - \frac{100\%}{T} \sum_{t=1}^{T} \frac{|x_t - \hat{x}_t|}{x_t} \]  
(38)

\[ \text{SR} = \frac{1}{T} \sum_{t=1}^{T} b_t \times 100\%, \quad b_t = \begin{cases} 1, & (\hat{x}_{t+1} - x_t)(x_{t+1} - x_t) \geq 0 \\ 0, & (\hat{x}_{t+1} - x_t)(x_{t+1} - x_t) < 0 \end{cases} \]  
(39)
In addition, the study used the standard DM test \cite{26} to further demonstrate the new model’s prediction advantage in terms of prediction accuracy. Assume that the initial $H_0$ is the two models’ prediction accuracy and is equal. The following are the statistics:

$$H_{dm} = \frac{\bar{o}}{\sqrt{\hat{R}_g/T}}; \quad \bar{o} = \frac{1}{T} \sum_{t=1}^{T} o_t, o_t = (x_t - x_{m,t})^2 - (x_t - x_{n,t})^2$$

$$\hat{R}_g = \lambda_0 + 2 \sum_{t=1}^{\infty} \lambda_t, \quad (\lambda_t = \text{cov}(o_{t+1}, o_t))$$

(40)

The variance of $o_t$ is denoted by $\lambda_0$, while the number of observations is represented by $T$. Moreover, $x_{m,t}$ and $x_{n,t}$ are the two models’ anticipated values for the $t$ period.

4. Empirical Analysis

This section clearly illustrates the proposed model’s prediction outcomes. In addition, the prediction results of other models are presented. Furthermore, eight assessment indices and tests are employed to demonstrate the effectiveness of the new model’s forecast predictions.

4.1. Empirical Results

Figures 7 and 8 show the SZA-2017 prediction results using four classical econometric models and seven artificial intelligence models, respectively. Similarly, the SZA-2019 forecast results are displayed in Figures 9 and 10. The chart shows that the trend of each single forecast model’s forecast series is generally the same as that of the actual series. Nevertheless, the fluctuation ranges in different periods are different. At the same time, the error curve shows that the errors of these models oscillate around 0. It can demonstrate that each model can gather data characteristics from various perspectives and develop its prediction advantages.

In addition, as shown by the scatter plot on the right, the predicted value of each model has a relatively positive linear relationship with the actual value. Furthermore, the convergence degree of the artificial intelligence model is somewhat higher when compared to the traditional measurement model.

Figure 7. The forecasting results, scatter plots, and error curves via classical measurement models (SZA-2017).
Figure 8. The forecasting results, scatter plots, and error curves via artificial intelligence models (SZA-2017).

Figure 9. The forecasting results, scatter plots, and error curves via classical measurement models (SZA-2019).
Because each model has different predictions of performance over time. Besides, the simple average combination model only takes a simple weighted average of the forecasted values of all individual models and cannot benefit from each model’s prediction benefits. As a result, we first use the Lasso regression model to screen the weights of some models to 0 and then apply the IOWA operator to build the optimal combination prediction model, thus obtaining the optimal prediction value.

The screening results of the Lasso regression model are shown in Table 2. It can be observed that almost all of the selected models are AI models. At the same time, the weights assigned to each model are different for distinct data sets, indicating that the training effects of models for different feature sequences are heterogeneous. In addition, Figures 11 and 12 exhibit the prediction results of the simple average prediction model and the proposed model for two different data sets. The prediction series of the constituted model is closer to the actual value than the simple average model, with more stable error fluctuations and stronger correlation.
Figure 11. The forecasting results, scatter plots, and error curves via combined forecasting models. (SZA-2017).

Figure 12. The forecasting results, scatter plots, and error curves via combined forecasting models. (SZA-2019).
4.2. Forecasting Effectiveness

4.2.1. Prediction Evaluation

This study examines eight evaluation indicators using the SZA-2017 and SZA-2019 data sets as examples. Tables 3–5 show the eight evaluation index values of all models. By analyzing the contents of the table, the following conclusions can be drawn:

(1) The model substituted in this study has a good prediction performance. The four error indicators of SZA-2017 are 0.2011, 1.2128, 0.1417, and 0.0242, and it has a UI value of 0.0143. These results are significantly smaller than other benchmark models. In addition, the SZA-2019 results are similar. In addition, considering the model’s generalization ability and prediction efficacy, the model’s IA, SR, and accuracy values are ideal, with 0.9674, 0.8621, and 0.9758, respectively, which are higher than other models. For SZA-2019, all other values are the best except SR (second only to GRNN). As a result, the proposed model has proven to be capable of accurate predictions.

(2) The Lasso regression model provides a robust model screening capacity for carbon price prediction. In this study, the Lasso regression model gives LSTM and LSSVM a heavier weight in SZA-2017, which corresponds to their prediction accuracy in this data set. The error indices of the LSTM model in SZA-2017 are 0.4586, 6.3087, and 0.3741 and a UI value of 0.0322, which are superior than other single prediction models. At the same time, its IA, SR, and accuracy values provide a significant benefit, indicating that it has improved generalization and prediction effectiveness. Furthermore, all of LSSVM’s index values are close to ideal. Likewise, the SZA-2019 analysis results lead to a similar conclusion. As a result, the Lasso regression model is helpful in further improving the constructed model’s forecasting accuracy.

(3) The simple combination model outperforms most single prediction models in terms of accuracy. The simple combination prediction model’s prediction errors in SZA-2017 are 0.5543, 9.2173, 0.4586, 0.0723, and it has a UI value of 0.0391. These values are much lower than the four traditional measurement models. It is, however, not superior to all artificial intelligence models. The values of the other three indices are 0.8958, 0.9277, and 0.7931, which are higher than traditional econometric models but lower than some AI models. Likewise, this finding can also be reached by looking at the SZA-2019 index value. Specifically, the underlying reason for this is that the basic average combination model weights all models equally and ignores the differences in prediction between models for various sequences, resulting in information redundancy or over-fitting, thus diminishing prediction ability to some extent.

(4) The prediction outcomes of artificial intelligence models and conventional econometric models both have their own benefits for various data sets. It is discovered that artificial intelligence models do not always perform better than conventional econometric models, and for various data sets, the metrics for each model’s accuracy, reliability, and generalizability of prediction findings fluctuate correspondingly. For instance, the RMSE of ARIMA is lower than that of BPNN, GRNN, NARX, and RNN for the SZA-2017 training set.

Table 3. Forecasting performance comparisons of classical measurement models.

<table>
<thead>
<tr>
<th>Data Sets</th>
<th>Algorithms</th>
<th>RMSE</th>
<th>SSE</th>
<th>MAE</th>
<th>MAPE</th>
<th>UI</th>
<th>IA</th>
<th>Accuracy</th>
<th>SR</th>
</tr>
</thead>
<tbody>
<tr>
<td>SZA-2017</td>
<td>ARIMA</td>
<td>0.6194</td>
<td>11.5107</td>
<td>0.539</td>
<td>0.0795</td>
<td>0.0442</td>
<td>0.9027</td>
<td>0.9205</td>
<td>0.4483</td>
</tr>
<tr>
<td></td>
<td>ES1</td>
<td>1.3184</td>
<td>52.1434</td>
<td>1.0977</td>
<td>0.1762</td>
<td>0.0922</td>
<td>0.676</td>
<td>0.8238</td>
<td>0.5172</td>
</tr>
<tr>
<td></td>
<td>ES2</td>
<td>0.8903</td>
<td>23.7781</td>
<td>0.7431</td>
<td>0.1108</td>
<td>0.0634</td>
<td>0.8628</td>
<td>0.8892</td>
<td>0.3103</td>
</tr>
<tr>
<td></td>
<td>GARCH</td>
<td>0.6294</td>
<td>11.8839</td>
<td>0.536</td>
<td>0.0791</td>
<td>0.0448</td>
<td>0.8991</td>
<td>0.9209</td>
<td>0.4483</td>
</tr>
<tr>
<td>SZA-2019</td>
<td>ARIMA</td>
<td>1.7926</td>
<td>96.3997</td>
<td>1.6175</td>
<td>0.1084</td>
<td>0.0532</td>
<td>0.981</td>
<td>0.8916</td>
<td>0.3793</td>
</tr>
<tr>
<td></td>
<td>ES1</td>
<td>4.1694</td>
<td>521.5287</td>
<td>3.6116</td>
<td>0.2593</td>
<td>0.1254</td>
<td>0.9335</td>
<td>0.7407</td>
<td>0.4138</td>
</tr>
<tr>
<td></td>
<td>ES2</td>
<td>2.224</td>
<td>148.3787</td>
<td>1.8649</td>
<td>0.1267</td>
<td>0.0651</td>
<td>0.9792</td>
<td>0.8733</td>
<td>0.5172</td>
</tr>
<tr>
<td></td>
<td>GARCH</td>
<td>2.2891</td>
<td>157.2005</td>
<td>1.5438</td>
<td>0.1119</td>
<td>0.0678</td>
<td>0.9792</td>
<td>0.8881</td>
<td>0.5862</td>
</tr>
</tbody>
</table>
Table 4. Forecasting performance comparisons of artificial intelligence models.

<table>
<thead>
<tr>
<th>Data Sets</th>
<th>Algorithms</th>
<th>Evaluation Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>SZA-2017</td>
<td>BPNNF</td>
<td>RMSE: 0.6981; SSE: 14.6202; MAE: 0.5663; MAPE: 0.0865; UI: 0.0498; IA: 0.8768; Accuracy: 0.9135; SR: 0.5862</td>
</tr>
<tr>
<td></td>
<td>GRNN</td>
<td>RMSE: 0.9633; SSE: 27.8411; MAE: 0.7014; MAPE: 0.1254; UI: 0.0668; IA: 0.7287; Accuracy: 0.8746; SR: 0.8276</td>
</tr>
<tr>
<td></td>
<td>LSTM</td>
<td>RMSE: 0.4586; SSE: 6.3087; MAE: 0.3741; MAPE: 0.0603; UI: 0.0322; IA: 0.9349; Accuracy: 0.9397; SR: 0.8621</td>
</tr>
<tr>
<td></td>
<td>NARX</td>
<td>RMSE: 1.4983; SSE: 67.3485; MAE: 1.2305; MAPE: 0.1876; UI: 0.0984; IA: 0.8273; Accuracy: 0.8124; SR: 0.8621</td>
</tr>
<tr>
<td></td>
<td>NARNN</td>
<td>RMSE: 0.6013; SSE: 10.8466; MAE: 0.5142; MAPE: 0.0756; UI: 0.0431; IA: 0.8888; Accuracy: 0.9244; SR: 0.4828</td>
</tr>
<tr>
<td></td>
<td>LSSVM</td>
<td>RMSE: 0.5198; SSE: 8.1066; MAE: 0.4838; MAPE: 0.0744; UI: 0.0381; IA: 0.9116; Accuracy: 0.9256; SR: 0.7931</td>
</tr>
<tr>
<td>SZA-2019</td>
<td>BPNNF</td>
<td>RMSE: 2.1472; SSE: 138.3081; MAE: 1.4842; MAPE: 0.0979; UI: 0.0626; IA: 0.9798; Accuracy: 0.9021; SR: 0.6552</td>
</tr>
<tr>
<td></td>
<td>GRNN</td>
<td>RMSE: 0.968; SSE: 28.1109; MAE: 0.8753; MAPE: 0.0528; UI: 0.0281; IA: 0.9912; Accuracy: 0.9472; SR: 0.9655</td>
</tr>
<tr>
<td></td>
<td>LSTM</td>
<td>RMSE: 2.0836; SSE: 130.2406; MAE: 1.7657; MAPE: 0.1225; UI: 0.0647; IA: 0.9767; Accuracy: 0.8775; SR: 0.6897</td>
</tr>
<tr>
<td></td>
<td>NARX</td>
<td>RMSE: 2.0892; SSE: 130.9463; MAE: 1.6523; MAPE: 0.1183; UI: 0.0606; IA: 0.9811; Accuracy: 0.8817; SR: 0.4828</td>
</tr>
<tr>
<td></td>
<td>NARNN</td>
<td>RMSE: 2.0149; SSE: 121.7979; MAE: 1.5881; MAPE: 0.1152; UI: 0.0583; IA: 0.979; Accuracy: 0.8848; SR: 0.7241</td>
</tr>
<tr>
<td></td>
<td>LSSVM</td>
<td>RMSE: 2.7319; SSE: 223.9011; MAE: 2.3501; MAPE: 0.1476; UI: 0.0814; IA: 0.9773; Accuracy: 0.8524; SR: 0.7586</td>
</tr>
<tr>
<td></td>
<td>RNN</td>
<td>RMSE: 2.908; SSE: 253.7017; MAE: 2.3475; MAPE: 0.1654; UI: 0.0846; IA: 0.9686; Accuracy: 0.8346; SR: 0.3793</td>
</tr>
</tbody>
</table>

Table 5. Forecasting performance comparisons of combined forecasting models.

<table>
<thead>
<tr>
<th>Data Sets</th>
<th>Algorithms</th>
<th>Evaluation Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>SZA-2017</td>
<td>CF</td>
<td>RMSE: 0.5543; SSE: 9.2173; MAE: 0.4586; MAPE: 0.0723; UI: 0.0391; IA: 0.8958; Accuracy: 0.9277; SR: 0.7931</td>
</tr>
<tr>
<td></td>
<td>BCF</td>
<td>RMSE: 0.2011; SSE: 1.2128; MAE: 0.1417; MAPE: 0.0242; UI: 0.0143; IA: 0.9674; Accuracy: 0.9758; SR: 0.8621</td>
</tr>
<tr>
<td>SZA-2019</td>
<td>CF</td>
<td>RMSE: 1.23; SSE: 45.3865; MAE: 1.0285; MAPE: 0.0727; UI: 0.0364; IA: 0.987; Accuracy: 0.9273; SR: 0.8966</td>
</tr>
<tr>
<td></td>
<td>BCF</td>
<td>RMSE: 0.3426; SSE: 3.521; MAE: 0.2842; MAPE: 0.0214; UI: 0.0102; IA: 0.9966; Accuracy: 0.9786; SR: 0.8966</td>
</tr>
</tbody>
</table>

4.2.2. Diebold–Mariano Test

The DM test is used to confirm the model’s prediction validity from a statistical standpoint. Tables 6–9 provide the inspection’s findings, where the above models are utilized as benchmark models for inspection. We can summarize the tables’ contents as follows:

(1) The prediction accuracy of the model proposed in this study is better than all other models selected when taking into account the findings of the DM test together with the predictive validity results in the preceding section. First of all, when compared to the conventional econometric prediction models, the model created for this study consistently rejected the original hypothesis at the 1% significance level. Therefore, the model performs better than each of the selected econometric models. In addition, the proposed model outperforms other artificial intelligence models relative to the benchmark AI model, except for GRNN, at the 1% or 5% significance level of predictive accuracy. Nevertheless, the proposed new model outperforms the GRNN model in terms of the eight criteria that were previously proposed to gauge the model’s predictive performance. Therefore, the investigation came to the conclusion that the new model outperforms all of the selected artificial intelligence models in terms of prediction accuracy. Furthermore, the prediction accuracy of the model created in this study was significantly better than the simple average combination model at the 1% level, with DM statistics of 3.5456 and 4.5468 for SZA-2017 and SZA-2019, respectively, indicating the importance of the study’s use of the Lasso model to select the optimal weights.

(2) Each type of single forecasting model has its benefits, while their prediction accuracy varies across different training sets. For instance, when using ARIMA as the baseline model to fit predictions to the SZA-2017 data, we discovered that at the 1% or 5% level, this model greatly outperformed the exponential smoothing prediction method, Holt’s exponential smoothing prediction method, NARX, and RNN. Their respective DM scores were −2.5585, −2.4201, −4.6137, and −3.1842. Nevertheless, when the model was used to
analyze SZA-2019, the outcomes drastically varied. When other individual models are used as benchmark models, the following pattern can be seen: the Holt’s exponential smoothing prediction method, LSTM, and NARX are no longer significantly less accurate than ARIMA, whereas the exponential smoothing prediction method, GARCH, LSSVM, and RNN are all significantly less accurate than ARIMA. It is clear that each forecasting model has a unique forecasting advantage due to its ability to capture the pertinent data in the carbon trading price series from various angles.

(3) The simple average combination model’s forecast accuracy was not always higher than that of the other single prediction models. According to the DM test findings, the prediction accuracy of the simple average combination model did not significantly exceed ARIMA, GRNN, NARNN, and LSSVM for the SZA-2017 training set and was even significantly lower than that of the LSTM model. In contrast, when SZA-2019 was used as the training set, the simple average combination model’s DM results at the 1% or 5% level outperformed those of ARIMA, the exponential smoothing prediction method, Holt’s exponential smoothing prediction method, NARX, NARNN, and RNN. This conclusion might be a result of the simple average combination model giving each single model the same weight while ignoring the interactions between the models and their suitability for the training dataset.

Table 6. DM test results of classical measurement models (SZA-2017).

<table>
<thead>
<tr>
<th>Target Model</th>
<th>ARIMA</th>
<th>Benchmark Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ES1</td>
</tr>
<tr>
<td>ES1</td>
<td>−2.5585</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(−0.0105)</td>
<td></td>
</tr>
<tr>
<td>ES2</td>
<td>−2.4201</td>
<td>1.9856</td>
</tr>
<tr>
<td></td>
<td>(−0.0155)</td>
<td>(−0.0471)</td>
</tr>
<tr>
<td>GARCH</td>
<td>−0.8166</td>
<td>2.5907</td>
</tr>
<tr>
<td></td>
<td>(−0.4142)</td>
<td>(−0.0096)</td>
</tr>
<tr>
<td>BPNNF</td>
<td>−1.0812</td>
<td>2.8282</td>
</tr>
<tr>
<td></td>
<td>(−0.2796)</td>
<td>(−0.0047)</td>
</tr>
<tr>
<td>GRNN</td>
<td>−0.8901</td>
<td>2.3232</td>
</tr>
<tr>
<td></td>
<td>(−0.3734)</td>
<td>(−0.0202)</td>
</tr>
<tr>
<td>LSTM</td>
<td>2.2642</td>
<td>2.9362</td>
</tr>
<tr>
<td></td>
<td>(−0.0236)</td>
<td>(−0.0033)</td>
</tr>
<tr>
<td>NARX</td>
<td>−4.6137</td>
<td>−0.5795</td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(−0.5622)</td>
</tr>
<tr>
<td>NARNN</td>
<td>0.6281</td>
<td>2.629</td>
</tr>
<tr>
<td></td>
<td>(−0.5300)</td>
<td>(−0.0086)</td>
</tr>
<tr>
<td>LSSVM</td>
<td>1.1887</td>
<td>2.6731</td>
</tr>
<tr>
<td></td>
<td>(−0.2346)</td>
<td>(−0.0075)</td>
</tr>
<tr>
<td>RNN</td>
<td>−3.1842</td>
<td>1.1200</td>
</tr>
<tr>
<td></td>
<td>(−0.0015)</td>
<td>(−0.2628)</td>
</tr>
<tr>
<td>CF</td>
<td>0.8795</td>
<td>3.0642</td>
</tr>
<tr>
<td></td>
<td>(−0.3791)</td>
<td>(−0.0022)</td>
</tr>
<tr>
<td>BCF</td>
<td>4.0330</td>
<td>3.1914</td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(−0.0014)</td>
</tr>
</tbody>
</table>
Table 7. DM test results of artificial intelligence models (SZA-2017).

<table>
<thead>
<tr>
<th>Target Model</th>
<th>Benchmark Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPNF</td>
<td>GRNN</td>
</tr>
<tr>
<td>GRNN</td>
<td>1.4343</td>
</tr>
<tr>
<td>LSTM</td>
<td>5.1172</td>
</tr>
<tr>
<td>NARX</td>
<td>5.1286</td>
</tr>
<tr>
<td>NARNN</td>
<td>5.1286</td>
</tr>
<tr>
<td>LSSVM</td>
<td>5.1286</td>
</tr>
<tr>
<td>RNN</td>
<td>5.1286</td>
</tr>
<tr>
<td>CF</td>
<td>5.1286</td>
</tr>
<tr>
<td>BCF</td>
<td>5.1286</td>
</tr>
</tbody>
</table>

Table 8. DM test results of classical measurement models (SZA-2019).

<table>
<thead>
<tr>
<th>Target Model</th>
<th>Benchmark Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA</td>
<td>ES1</td>
</tr>
<tr>
<td>ESI</td>
<td>3.5328</td>
</tr>
<tr>
<td>E2</td>
<td>0.1155</td>
</tr>
<tr>
<td>GARCH</td>
<td>0.0664</td>
</tr>
<tr>
<td>RNN</td>
<td>0.0712</td>
</tr>
<tr>
<td>NARX</td>
<td>0.0712</td>
</tr>
<tr>
<td>NARNN</td>
<td>0.0712</td>
</tr>
<tr>
<td>LSSVM</td>
<td>0.0712</td>
</tr>
<tr>
<td>RNN</td>
<td>0.0712</td>
</tr>
<tr>
<td>CF</td>
<td>0.0712</td>
</tr>
<tr>
<td>BCF</td>
<td>0.0712</td>
</tr>
</tbody>
</table>
Table 9. DM test results of artificial intelligence models (SZA-2019).

<table>
<thead>
<tr>
<th>Target Model</th>
<th>Benchmark Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>BPNF</td>
<td>GRNN</td>
</tr>
<tr>
<td>GRNN</td>
<td>1.771</td>
</tr>
<tr>
<td>LSTM</td>
<td>0.176</td>
</tr>
<tr>
<td>NARX</td>
<td>0.1416</td>
</tr>
<tr>
<td>NARNN</td>
<td>0.2353</td>
</tr>
<tr>
<td>LSSVM</td>
<td>0.9231</td>
</tr>
<tr>
<td>RNN</td>
<td>0.3559</td>
</tr>
<tr>
<td>CF</td>
<td>1.5538</td>
</tr>
<tr>
<td>BCF</td>
<td>2.0866</td>
</tr>
</tbody>
</table>

5. Conclusions and Discussion

Forecasting an accurate carbon price trend is crucial for the growth of the carbon emission trading market. Given the current carbon price forecasting constraints, an optimal combination carbon price forecasting model based on the Lasso regression model is developed. First, 11 single forecasting models are chosen to forecast the relevant data sets. After that, the Lasso regression model is employed to reduce the dimensionality. Thus, the weights of some models are set to 0. Moreover, the best combination prediction model based on the IOWA operator is used to forecast the remaining models. As a result, the optimal weight can be calculated to obtain the final prediction value. In addition, using the Shenzhen SZA-2017 and SZA-2019 data sets as samples, eight evaluation indices and the DM test are applied to assess the model's prediction effectiveness from various perspectives. The following breakthroughs and strengths of this paradigm have been identified through empirical and extensive analysis:

(1) Compared to existing benchmark models, the suggested model outperforms them in terms of prediction accuracy, generalization ability, and effectiveness. It introduces a novel strategy for predicting carbon prices.

(2) For the first time, the Lasso regression model is used in carbon price prediction, resolving the challenge of selecting a single prediction model in a combination prediction model.

(3) The simple average combination model’s suboptimal properties are empirically tested, and then the ideal weight is established.

In addition, the study also has some academic significance and practical benefits. On the one hand, the study builds a combined model for predicting carbon trading prices based on Lasso regression and optimal integration, as well as forecasts two different types of carbon trading prices in the Shenzhen carbon trading market, enhancing theoretical research on carbon trading price forecasting and the carbon trading market while also furthering forecasting results and their accuracy. Nevertheless, this study has significant practical implications for the policy making of government agencies, the strategic positioning of emission control businesses, and the decision making of investors. First and foremost, since China’s carbon trading market is currently in a stage of gradual improvement, the accurate price prediction of carbon trading is conducive to a deeper exploration of the intricate patterns of carbon trading prices, providing a reference for government departments to formulate policies related to the carbon trading market and accelerating the establishment of an effective and complete carbon price trading system, further improving the market.
regulatory mechanism. The study will also assist emission control businesses in predicting the trajectory of carbon trading prices and allocating resources more sensibly and effectively to cut costs and achieve sustainable development. The study can also give investors market information estimates, which can result in more useful investment and financing guidance, given that carbon emission rights are also a financial asset.

Nevertheless, there are still certain restrictions that need to be addressed. First, all of the forecasts in this research are based on historical carbon price data but do not take into account unstructured data, causing the prediction results to lag. More research is needed to determine how to mix this type of data with the combined forecasting model. Furthermore, the combination prediction model used combines traditional measurement methods with artificial intelligence methodologies. However, the decomposition and integration models are not taken into account. As a result, in the future, it may be essential to combine these methodologies with the optimal combination forecasting model to create a more general and robust forecasting model. Ultimately, to optimize its utility, the suggested model can be applied to the prediction of business cycles, economic growth, policy formulation, stock price forecasting, and other relevant domains.

Author Contributions: Funding acquisition, Y.L.; Validation, Y.L., X.W. and N.S.; Methodology, R.Y.; Writing, R.Y.; Editing, R.Y., X.W., J.Z. and N.S.; Review, X.W. and N.S.; Model design J.Z.; Formal analysis, J.Z.; Supervision, J.Z.; Visualization, N.S. All authors have read and agreed to the published version of the manuscript.

Funding: The work was supported by the National Natural Science Foundation of China (Nos. 72001001, 72103127) and the Anhui Provincial Natural Science Foundation (No. 2008085QG334).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available on request from the corresponding authors.

Conflicts of Interest: The authors declared that they have no conflicts of interest to report in this work.

References
8. Zhu, B.; Wei, Y. Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology. Omega 2013, 41, 517–524. [CrossRef]


19. Liu, Z.; Huang, S. Carbon option price forecasting based on modified fractional Brownian motion optimized by GARCH model in carbon emission trading. N. Am. J. Econ. Financ. 2021, 55, 1062–9408. [CrossRef]


**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.