

Article

A Novel Extended Higher-Order Moment Multi-Factor Framework for Forecasting the Carbon Price: Testing on the Multilayer Long Short-Term Memory Network

Po Yun ^{1,*}, Chen Zhang ^{1,2,*}, Yaqi Wu ¹, Xianzi Yang ¹ and Zulfiqar Ali Wagan ³

¹ School of Management, Hefei University of Technology, Hefei 230009, China; yq111099@mail.hfut.edu.cn (Y.W.); xianziyang@mail.hfut.edu.cn (X.Y.)

² Key Laboratory of Process Optimization and Intelligent Decision-Making, Ministry of Education, Hefei 230009, China

³ Department of Education and Literacy, Government of Sindh, Hyderabad 70060, Tando Jam, Pakistan; zulfiqarsch@mail.hfut.edu.cn

* Correspondence: yunpo2010@mail.hfut.edu.cn (P.Y.); sm.zhangchen@hfut.edu.cn (C.Z.)

Received: 15 February 2020; Accepted: 27 February 2020; Published: 2 March 2020



Abstract: Predicting the carbon price accurately can not only promote the sustainability of the carbon market and the price driving mechanism of carbon emissions, but can also help investors avoid market risks and increase returns. However, previous research has only focused on the low-order moment perspective of the returns for predicting the carbon price, while ignoring the shock of extreme events and market asymmetry originating from its pricing factor markets. In this paper, a novel extended higher-order moment multi-factor framework (EHM-APT) was formed to improve the prediction and to capture the driving mechanism of the carbon price. Furthermore, a multi-layer and multi-variable Long Short-Term Memory Network (Multi-LSTM) model was constructed so that the parameters and structure could be determined experimentally for testing the performance of the proposed framework. The results show that the pricing framework considers the shock of extreme events and market asymmetry and can improve the prediction compared with a framework that does not consider the shock of higher-order moment terms. Additionally, the Multi-LSTM model is more competitive for prediction than other benchmark models. This conclusion proves the rationality and accuracy of the proposed framework. The application of the pricing framework encourages investors and financial institutions to pay more attention to the pricing factor of extreme events and market asymmetry for accurate price prediction and investment analysis.

Keywords: co-skewness; co-kurtosis; higher-order moment; carbon price forecasting; Multi-factor; Multi-LSTM

1. Introduction

Upon signing the Kyoto protocol, the carbon market was formally established in 2005 in an attempt to reduce global greenhouse gas. The carbon market defines virtual carbon emission rights as scarce valuable assets, gives them commodity attributes, and realizes the target of resource allocation and emission reduction through market transaction among reduction entities. The signing of the Paris Agreement in December 2015 further highlights the carbon market's capital allocation for achieving emission reduction on a global scale [1]. As a core issue of the carbon market mechanism, accurate forecasting of the carbon price can develop an efficient carbon pricing mechanism, and also help investors to avoid market risks and to increase returns. As an emerging policy-based artificial market,

the carbon market is characterized by strong sensitivity to policy shocks, especially carbon dioxide (CO₂) reduction policies and carbon quota policies [2]. Therefore, as for the certain pricing framework, the prediction of the carbon price should not only follow the basic pricing method of general financial assets, but also reflect the special driving mechanism of the carbon price.

Further evidence has verified that the shock of policy turbulence (like the implementation of carbon tax) and uncertainty in the energy and financial markets can easily affect the carbon price through financial channels [3]. In other words, the shock from the energy and financial markets will also have an impact on the carbon price forecast and price fluctuation. Consequently, the carbon price is vulnerable to national strategies, government regulations, international agreements, and other policy factors, as well as asymmetric influences from market participants' behaviors that originate from its influential markets. Thus, it is worth exploring the carbon price driving mechanism, which reflects the impact of external shock and market asymmetry.

However, the common foundation of existing research only focuses on the low-order moment perspective of the returns (market return and variance), while ignoring the shock of extreme events and market asymmetry originating from its influential markets from the framework of higher-order moment (market skewness and kurtosis). That is, these ignored factors are not included in the analysis framework of the pricing mechanism. Actually, market skewness and kurtosis have proven to be valuable indicators that characterize market asymmetry information and market extreme factors for explaining portfolio returns [4]. This kind of defect makes the accuracy of the existing carbon price prediction questionable.

To overcome this weakness, the multivariate skew-GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model, which incorporates the third-order moment of financial assets, was introduced to explain finance return and the asymmetric relation among some European financial markets [5]. However, the multivariate skew-GARCH model has more parameters to be estimated, and needs certain assumptions on the tail distribution of the residual. Furthermore, Fry et al. [6] developed a binary higher-order moment Capital Asset Pricing Model (CAPM) model, which considers the shock of co-skewness and co-kurtosis to represent the risk of market asymmetry and extreme event on the return of the portfolio. As a result, the higher-order moment CAPM model provides a reasonable framework to study the pricing of carbon assets under the impact of external events and information asymmetry. As we know, policy factors and information asymmetry are important external factors affecting the carbon price. While many scholars use Extreme Value Theory (EVT) to examine extreme risk shocks in financial markets [7], the EVT model ignores the extreme shock from other relevant markets and only considers the extreme impact generated by the financial market itself. Thus, the higher-order moment CAPM model is superior to EVT-based models in describing the asset prices affected by the extreme event shocks.

However, the higher-order moment CAPM model mentioned above is limited in its ability to describe the relationship between two assets, and requires that the asset returns obey the binary standard normal distribution, which makes it difficult to capture the non-linear mapping relationship between the carbon price and the multiple pricing factors.

The aim of this article is to construct a novel extended higher-order moment multi-factor framework for forecasting the carbon price, in which the framework is in line with the non-linear and peak thick tail characteristics of the carbon price. Furthermore, this paper relaxes the framework of higher-order moment CAPM mentioned above to multivariate factors, and brings the higher-order moment terms that indicate the shock of market asymmetry and extreme events originating from its influential markets into the carbon pricing framework. Thus, based on the arbitrage pricing theory (APT), an extended higher-order moment multi-factor framework (EHM-APT) that conforms to the special driving mechanism of the carbon price is formed. The above research design is to compensate for the inaccurate pricing caused by neglecting the pricing factors of extreme events and market asymmetry information. The proposed framework is convinced that the carbon price can be theoretically explained by the higher-order moment terms of its influential markets. Further, this article uses a machine learning

method to empirically evaluate the accuracy and performance of the framework. Specifically, this paper utilizes the Long Short-Term Memory network with the designed multi-layer and multi-variable structure (Multi-LSTM) to predict the carbon price and capture the non-linear mapping relationship between the carbon price and its pricing factors.

The conclusion of this article summarizes that the proposed EHM-APT framework can significantly improve the prediction of the carbon price. That is to say, the co-skewness and co-kurtosis, which represent the market asymmetry and extreme events, which stem from its pricing factor markets, are indispensable factors for predicting and fitting the carbon price. Therefore, it is valuable to allocate and arrange funds for carbon investors by taking market extreme factors and market asymmetric information as indispensable pricing factors. The conclusion of this article and the proposed pricing framework will help investors to predict and analyze the carbon price both effectively and accurately.

The remaining structure of this paper is as follows: Section 2 reviews the related literature. Section 3 introduces the methodology of this paper. Section 4 presents the experiment analysis, contains the descriptive statistics and the design of the Multi-LSTM model, and offers the results and the discussion. Section 5 concludes this paper and puts forward ideas for further research.

2. Literature Review

The growing academic literature on carbon price forecasting has focused on two major perspectives: One is the traditional econometric models, the other is the artificial intelligence models.

For traditional econometric models, the GARCH model has potential advantages in capturing the non-linear characteristics of the carbon price. Chevallier [8] proposed that the asymmetric threshold GARCH model can effectively capture the characteristics of the carbon future price and improve prediction accuracy. Byun et al. [9] put forward that GARCH-type models are superior to the k-nearest neighbor model for predicting the carbon price. Koop and Tole [10] created the dynamic model averaging (DMA) method to forecast the carbon price, and the results showed a high prediction accuracy compared with the Bayesian model and the time-varying parameter regression model. Sanin et al. [11] demonstrated that the integration of Autoregressive Moving Average model (ARMAX) and the GARCH model, when using a time-varying jump process, can accurately predict the carbon price compared with a standard ARMAX–GARCH framework. Benz et al. [12] concluded that the Markov regime-switching GARCH model outperforms other Markov regime-switching or simple GARCH models in forecasting the carbon price. In addition, Chevallier [13] maintained that a nonparametric method of predicting the carbon price can reduce the prediction error by almost 15% compared with linear autoregression models. Based on the integration of the threshold dynamic conditional correlation (DCC) GARCH model and the full Baba, Engle, Kraft, and Kroner (BEKK) GARCH model, the volatility spillover from the price decrease of the energy market may have a stronger shock on carbon price, and the GARCH-based model can better reveal the price volatility mechanism of carbon assets based on the analysis of volatility spillover [14,15].

However, the applicability of traditional econometric models usually requires that the returns follow a strict hypothesis, such as normal distribution at the tail of carbon returns [16]. As a matter of fact, the carbon return has significant non-normal and non-linear characteristics, and asset returns exhibit a peak and thick tail feature [8], which results in traditional pricing methods failing to predict the carbon price accurately.

Being capable of mapping non-linear functions and not needing to consider the tail distribution of return series, artificial intelligence methods include artificial neural network (ANN), support vector machine (SVM), least squares support vector machine (LSSVM), multilayer perceptron (MLP), and Hybrid Fuzzy Neural Network (HFNN), which are superior to traditional econometric models in solving forecasting issues [17,18]. To determine the parameters of the LSSVM model, Zhu et al. [19] proposed an integration approach of a group method of data handling (GMDH), particle swarm optimization (PSO), and the LSSVM model for carbon price prediction, in which the prediction accuracy is higher than that of ANN and the autoregressive integrated moving average (ARIMA) model.

Fan et al. [20] established an MLP-ANN prediction model, and the results showed that the proposed model has good performance compared with the models of ARIMA, ANN, and LSSVM. Recently, research has found that an integrated model that combines the EMD (Empirical model decomposition) method with ANN and LSSVM can achieve better performance for forecasting the carbon price than that of the EMD method alone [21]. Additionally, Atsalakis [22] proposed a computational intelligence-based model with a novel hybrid neuro-fuzzy controller for forecasting the carbon price, which obtained a higher accuracy. Zhu et al. [23] combined variational mode decomposition (VMD) and spiking neural networks (SNNs) to improve forecasting accuracy and reliability.

There are two defects in the above research. Firstly, the foundation of these studies may ignore the shock of the higher-order moment attribute (skewness-kurtosis) of the pricing factors on the carbon price. The common theoretical basis of the above models is exploring the linear or non-linear regression relationship between the carbon price and its pricing factors under the multi-factor framework, which is a manner of the low-order moment of return. Correspondingly, the previous studies imply the hypothesis that the shock of extreme events or information asymmetry that stem from its pricing factor markets will not affect carbon price. Secondly, the forecasting models mentioned in previous research are flawed in their ability to predict carbon price accurately. Because the traditional statistical and econometric models require that the return follows a strict hypothesis, artificial intelligence methods are easily able to fall into local minima for solving forecasting issues and have difficulty in achieving a globally optimal situation.

Therefore, to remedy the shortcomings of the existing research, the contributions of this paper are as follows: firstly, this article relaxes the framework of binary higher-order moment CAPM to multivariate factors. That is, to bring the higher-order moment terms that indicate the shock of market asymmetry and extreme events into the carbon pricing framework. As for the framework, a complex non-linear and non-structural relationship between the carbon price and its pricing factors cannot be ignored. Therefore, the second innovation is the creation of a model of Multi-LSTM with the designed multi-layer and multi-variable structure to capture the non-linear mapping relationship between the carbon price and its pricing factors. The reasons for selecting the Multi-LSTM network are as follows: on the one hand, the Multi-LSTM has an advantage in handling time series data with non-linear and complex relationships, which is consistent with the time lag characteristic of the carbon price sequence. On the other hand, a remarkable feature of Multi-LSTM in the training process is that the model adjusts adaptively for obtaining optimal parameters and structure, and can also avoid falling into the local optimal solution [24], rather than the other optimization models such as the genetic algorithm and the backpropagation algorithm used in the process of neural network training [25,26].

3. Methodology

In this section, we theoretically explain the construction process of the novel extended higher-order moment multi-factor pricing framework. Based on this, the Multi-LSTM network for fitting the pricing framework is introduced.

3.1. A Novel Extended Higher-Order Moment Multi-Factor Framework (EHM-APT)

According to the higher-order moment CAPM model proposed by Hwang et al. [27], the return of financial assets is affected not only by systemic risk, but also by irrational behavior and extreme external shock represented by co-skewness and co-kurtosis. Therefore, the higher-order moment CAPM model can reveal more price characteristics than the traditional CAPM model.

The higher-order moment CAPM model, according to Hwang et al. [27], is as follows:

$$E(R_i) - R_f = a_1\beta_{im} + a_2\gamma_{im} + a_3\delta_{im} \quad (1)$$

and

$$\beta_{im} = \frac{E[(r_{i,t} - E(r_{i,t})) \times (r_{m,t} - E(r_{m,t}))]}{E(r_i) \times E(r_m)} \quad (2)$$

$$\gamma_{im} = \frac{E[(r_{i,t} - E(r_{i,t})) \times (r_{m,t} - E(r_{m,t}))^2]}{E(r_i) \times E(r_m)^2} \quad (3)$$

$$\delta_{im} = \frac{E[(r_{i,t} - E(r_{i,t})) \times (r_{m,t} - E(r_{m,t}))^3]}{E(r_i) \times E(r_m)^3} \quad (4)$$

where $E(R_i) - R_f$ is the excess return of the asset portfolio, a_1, a_2, a_3 represents the risk premium coefficient of $\beta_{im}, \gamma_{im}, \delta_{im}$ respectively. β_{im} is the coefficient of co-variance and represents the shock of the first-order centered moment (return) of asset portfolio m on the first-order centered moment (return) of asset i, γ_{im} is the coefficient of co-skewness and represents the shock of the second-order centered moment (variance) of asset portfolio m on the first-order centered moment (return) of asset i, and δ_{im} is the coefficient of co-kurtosis and represents the shock of the third-order centered moment (skewness) of asset portfolio m on the first-order centered moment (return) of asset i. The co-skewness reflects the information asymmetry of market investment, and the negative skewness results in the probability of return decline exceeding the probability of rise, thus increasing the losses for investors [6]. Correspondingly, higher co-kurtosis increases the probability of extreme events, which reflects the impact of external events on market investment.

For examining the shock of higher-order moment terms on the portfolio return, Fry et al. [6] developed an extended higher-order moment CAPM model by introducing co-skewness and co-kurtosis into the traditional CAPM framework. The suggested model, which incorporates bivariate normal distribution under the shock of co-skewness and co-kurtosis, is shown as

$$f_{\text{coskness}}(r_{1,t}, r_{2,t}) = \exp\left(-\frac{1}{2}\left(\frac{1}{1-\rho^2}\right)\left(\left(\frac{r_{1,t}-\mu_1}{\sigma_1}\right)^2 + \left(\frac{r_{2,t}-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{r_{1,t}-\mu_1}{\sigma_1}\right)\left(\frac{r_{2,t}-\mu_2}{\sigma_2}\right)\right)\right) + \theta_{12}\left(\frac{r_{1,t}-\mu_1}{\sigma_1}\right)^1\left(\frac{r_{2,t}-\mu_2}{\sigma_2}\right)^2 - \eta \quad (5)$$

$$f_{\text{cokurtosis}}(r_{1,t}, r_{2,t}) = \exp\left(-\frac{1}{2}\left(\frac{1}{1-\rho^2}\right)\left(\left(\frac{r_{1,t}-\mu_1}{\sigma_1}\right)^2 + \left(\frac{r_{2,t}-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{r_{1,t}-\mu_1}{\sigma_1}\right)\left(\frac{r_{2,t}-\mu_2}{\sigma_2}\right)\right)\right) + \theta_{13}\left(\frac{r_{1,t}-\mu_1}{\sigma_1}\right)^1\left(\frac{r_{2,t}-\mu_2}{\sigma_2}\right)^3 - \eta \quad (6)$$

where $f_{\text{coskness}}(r_{1,t}, r_{2,t})$ and $f_{\text{cokurtosis}}(r_{1,t}, r_{2,t})$ represents the portfolio return considering the impact of co-skewness and co-kurtosis, respectively, θ_{12} represents the impact coefficient of co-skewness, that is, the shock of the second-order centered moment (variance) of asset 2 on the first-order centered moment (return) of asset 1, θ_{13} denotes the impact coefficient of co-kurtosis, that is, the shock of the third-order centered moment (skewness) of asset 2 on the first-order centered moment (return) of asset 1, $r_{1,t}, \mu_1, \sigma_1$ and $r_{2,t}, \mu_2, \sigma_2$ represent the return, mean, and variance of assets 1 and 2, respectively, ρ refers to the correlation coefficient between assets 1 and 2, and η indicates the residual.

However, the above extended higher-order moment CAPM model is only applicable to the pricing of binary assets, and the assumption of the binary standard normal distribution is required. To overcome this limitation, this paper proposes an extended higher-order moment multi-factor framework by introducing the co-skewness and co-kurtosis of multivariate pricing factors into the framework. The novel pricing foundation, which considers the shock of market asymmetry and extreme events, is shown as

$$f(r_{0,t}) = \alpha_1 r_{1,t} + \alpha_2 r_{2,t} + \alpha_3 r_{3,t} + \dots + \alpha_n r_{n,t} + \sum_{i=1}^n \theta_i \left(\frac{r_{0,t}-\mu_0}{\sigma_0}\right)^1 \left(\frac{r_{i,t}-\mu_i}{\sigma_i}\right)^2 + \sum_{i=1}^n \delta_i \left(\frac{r_{0,t}-\mu_0}{\sigma_0}\right)^1 \left(\frac{r_{i,t}-\mu_i}{\sigma_i}\right)^3 + \eta_{i,t} \quad (7)$$

where

$$\eta_{i,t} = \ln \int \int \dots \int \exp[\alpha_1 r_{1,t} + \alpha_2 r_{2,t} + \alpha_3 r_{3,t} + \dots + \alpha_n r_{n,t} + \sum_{i=1}^n \theta_i \left(\frac{r_{0,t}-\mu_0}{\sigma_0}\right)^1 \left(\frac{r_{i,t}-\mu_i}{\sigma_i}\right)^2 + \sum_{i=1}^n \delta_i \left(\frac{r_{0,t}-\mu_0}{\sigma_0}\right)^1 \left(\frac{r_{i,t}-\mu_i}{\sigma_i}\right)^3] dr_1 dr_2 \dots dr_n \quad (8)$$

where $f(r_{0,t})$ represents the carbon return, $\alpha_1, \alpha_2, \dots, \alpha_n$ represents the impact coefficient of the pricing factors, $r_{0,t}, \mu_0, \sigma_0$ and $r_{i,t}, \mu_i, \sigma_i$ represent the return, mean, and variance of the carbon market and its pricing factors, respectively, θ_i denotes the impact coefficient of the second-order centered moment (variance) of carbon pricing factors on the first-order centered moment (return) of carbon return, that is, the shock of market asymmetry, and δ_i denotes the impact coefficient of the third-order centered moment (skewness) of carbon pricing factors on the first-order centered moment (return) of carbon return, that is, the shock of extreme events.

3.2. A Multi-Layer and Multi-Variable LSTM (Multi-LSTM) Model for Predicting the Carbon Price

This paper constructs a multi-layer and multi-variable LSTM model based on the research of Hochreiter et al. [24] for predicting the carbon price and investigating the performance of the proposed EHM-APT framework.

The LSTM network consists of an input layer, an output layer, and a hidden layer. Remarkably, the hidden layer of the LSTM is no longer a common neuron compared with the traditional Recurrent Neural Network (RNN), but rather an LSTM cell with special memory function. As the core of LSTM, the cell can eliminate or enhance the transmission of input data to the cell by controlling the structure of the designed gate. That is, the forget gate, the input gate, and the output gate. Furthermore, the Multi-LSTM model contains more cells and hidden layers based on single layers. The Multi-LSTM with a multi-layer chain structure and recursive network features can extract and mine the feature of input data, thus significantly improving the ability of learning and training.

During the training of LSTM, the three gates of LSTM receive two kinds of external information at time t , that is, X_t represents the input of the current state and H_{t-1} denotes the output of the hidden layer at the previous state. In addition, each gate receives internal information, that is, the memory unit of C_{t-1} at the previous state. Among the three gates, the forget gate determines the discarded information in the cell, which is equivalent to decay of information in each dimension [24]. The input gate determines the type of updated and input information in the cell. In this process, a sigmoid function is responsible for updating the information and the tanh function creates a new candidate vector of C_t at the current cell state. The output gate determines the output of the current cell through the sigmoid function, while the tanh function is used to convert the current output into a value between -1 and 1 , and then the final output of LSTM is obtained after multiplying the output of the above two functions. The training structure of the Long Short-Term Memory network is shown in Figure 1.

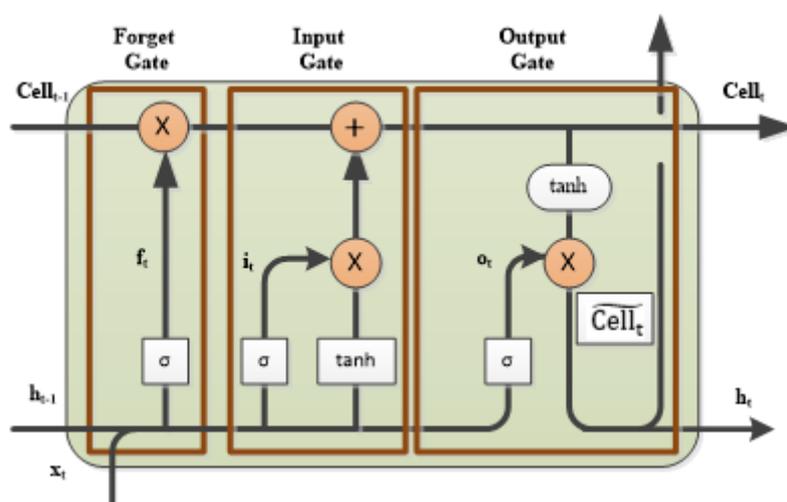


Figure 1. The training structure of the Long Short-Term Memory (LSTM) network.

The information filtering of the forget gate is shown as

$$f_t = \sigma(W_f \times [h_{t-1}, x_t] + b_f) \quad (9)$$

The information updating of the input gate is shown as

$$i_t = \sigma(W_i \times [h_{t-1}, x_t] + b_i) \quad (10)$$

$$\tilde{C}_t = \tanh(W_C \times [h_{t-1}, x_t] + b_C) \quad (11)$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (12)$$

The information screening of the output gate is shown as

$$o_t = \sigma(W_o \times [h_{t-1}, x_t] + b_o) \quad (13)$$

$$h_t = o_t \times \tanh(C_t) \quad (14)$$

where \otimes means the Kronecker product and \oplus means the XOR logic operation. The information updating vector, the candidate vector, and the update vector of the input gate at the current cell are represented by i_t , \tilde{C}_t , and C_t . h_t refers to the final output of the LSTM network. W_f , W_i , W_C , and W_o represent the weight vectors. b_f , b_i , b_C , and b_o show the bias of the training process and σ is the sigmoid function.

3.3. Evaluation Criteria of Multi-LSTM for Determining the Performance of the Proposed Framework

This paper conducts the following criteria to evaluate the parameters and performance of the Multi-LSTM model. The five criteria are

$$RMSE = \sqrt{\frac{\sum_{i=1}^T (y_i - \hat{y}_i)^2}{T}} \quad (15)$$

$$MAE = \frac{1}{T} \sum_{i=1}^T |y_i - \hat{y}_i| \quad (16)$$

$$MAPE = \frac{1}{T} \sum_{i=1}^T \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (17)$$

$$DA = \frac{1}{T} \sum_{i=1}^T a_i \text{ where } a_i = \begin{cases} 1, & \text{if } (y_{i+1} - y_i) \times (\hat{y}_{i+1} - \hat{y}_i) > 0 \\ 0, & \text{otherwise} \end{cases} \quad (18)$$

$$CORR = \frac{E(Y\hat{Y}) - E(Y) \times E(\hat{Y})}{\sqrt{E(Y^2) - E^2(Y)} \times \sqrt{E(\hat{Y}^2) - E^2(\hat{Y})}} \quad (19)$$

where $Y = \{y_1, y_2, \dots, y_T\}$ represents the time series of actual carbon return, $\hat{Y} = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_T\}$ means the series of predicted carbon return, and T denotes the observations.

Furthermore, we chose the criteria of Root-Mean-Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) because they are utilized in model performance judgment and are often used in the existing literature. These indicators are all real numbers and are between 0 and 1. If their values are smaller, then the closer the true value to the predicted value, and thus the better the prediction performance of the model. As for the investors, the criterion of direction accuracy (DA) means the probability that the market trend is consistent with the correctly predicted direction. Investors generally pay more attention to the probability of the market's predicted value deviating from the true value, as the right investment direction can help investors make more valuable decisions. If the value of DA is larger, it means the predicted carbon return is closer to the psychological expectation of investors. CORR represents the Pearson correlation coefficient, which is used to measure

the correlation between the true carbon return and the predicted return. The value of CORR is between 0 and 1. The closer the CORR value is to 1, the stronger the correlation, and vice versa.

4. Experiment Analysis and Discussion

4.1. The Data and Basic Statistical Analysis

The pricing factors of the carbon market include the prices of carbon homogeneous products, capital market products, and energy market products [13]. Therefore, this paper selects the EUAF (European Union Allowance Future) as the representative product of the carbon market for its large volume in EU ETS (European Union Emissions Trading System). Additionally, we chose the EUAS (European Union Allowance Spot) as the variable of carbon homogeneous products, took the DJIA (Dow Jones Industrial Average) and EURUSD as the typical products of capital market, and selected coal, crude oil, and natural gas as the representations of the energy market for their larger consumption. The data of EUAF and EUAS were sourced from the database of Intercontinental Exchange (ICE), and the other data were sourced from the Wind database. The research samples ranged from the period of 2 June 2009 to 8 May 2019, and there was a total of 2521 observations by eliminating the samples with missing information or time inconsistency. The return is expressed as: r_t , and $r_t = 100 \times (\ln P_t - \ln P_{t-1})$, where P_t denotes the product's settlement price.

We can draw the following conclusions from the descriptive statistics, as shown in Table 1: Firstly, all of the return series have good stationarity and are significant at the 1% level. Secondly, the mean difference of each return is small, but the standard deviation is big. As an example, the products of EUAF, EUAS, and gas show larger variance statistics, while that of EURUSD has a smaller variance. Thirdly, all of the return series show negative skewness, while the kurtosis values of DJIA and gas are the largest. This result indicates that the probability of return decline is greater than that of rise, and these markets are more vulnerable to the shock of extreme events. Furthermore, this paper also discloses the co-skewness and co-kurtosis statistics of each sample market. Negative co-skewness indicates information asymmetry from the pricing factor markets, which can easily increase the probability of a decline in carbon return. Positive co-kurtosis, especially the larger co-kurtosis presented by EUAS and DJIA, indicates that the external events suffered by these two markets will spread to the carbon market in a greater probability.

Table 1. Descriptive statistics.

	EUAF	EUAS	DJIA	EURUSD	Coal	Oil	Gas
ADF	−37.846 ***	−38.472 ***	−43.7 ***	−52.285***	−53.306 ***	−47.82 ***	−32.362 ***
Mean	0.0197	0.023	0.0429	−0.0097	0.0013	−0.0129	−0.0211
Std.Dev.	3.1025	3.2537	1.3054	0.5756	1.9056	1.2781	3.9581
Skewness	−0.2114	−0.9289	−0.9225	−0.0826	−0.0013	−1.4089	−0.0362
Kurtosis	19.1693	20.5059	353.4518	4.5682	5.8416	42.2814	69.4135
Coskewness12		−0.9686	−0.1506	−0.0647	−0.0462	−0.1343	−0.0146
Cokurtosis13		15.484	10.0999	0.2109	0.3501	0.3606	0.5036

Note: *** denotes the statistical significance at the 1% level. Coskewness12 and Cokurtosis13 mean the co-skewness and co-kurtosis coefficients between EUAF and the other market products, and the value is calculated by $E(r_{x,t} - E(r_{x,t}))(r_{y,t} - E(r_{x,t}))^2 / \sigma_x \sigma_y^2$ and $E(r_{x,t} - E(r_{x,t}))(r_{y,t} - E(r_{x,t}))^3 / \sigma_x \sigma_y^3$, respectively, where $r_{x,t}$ and $r_{y,t}$ denote the return of EUAF and the other market products.

Furthermore, in the joint distribution of EUAF and DJIA, EURUSD, coal has a large kurtosis, as shown in Figure 2. This indicates that the portfolios are more vulnerable to the shock of extreme events. This result provides further strong evidence for studying the shock of pricing factor markets on the carbon price from the perspective of higher-order moment.

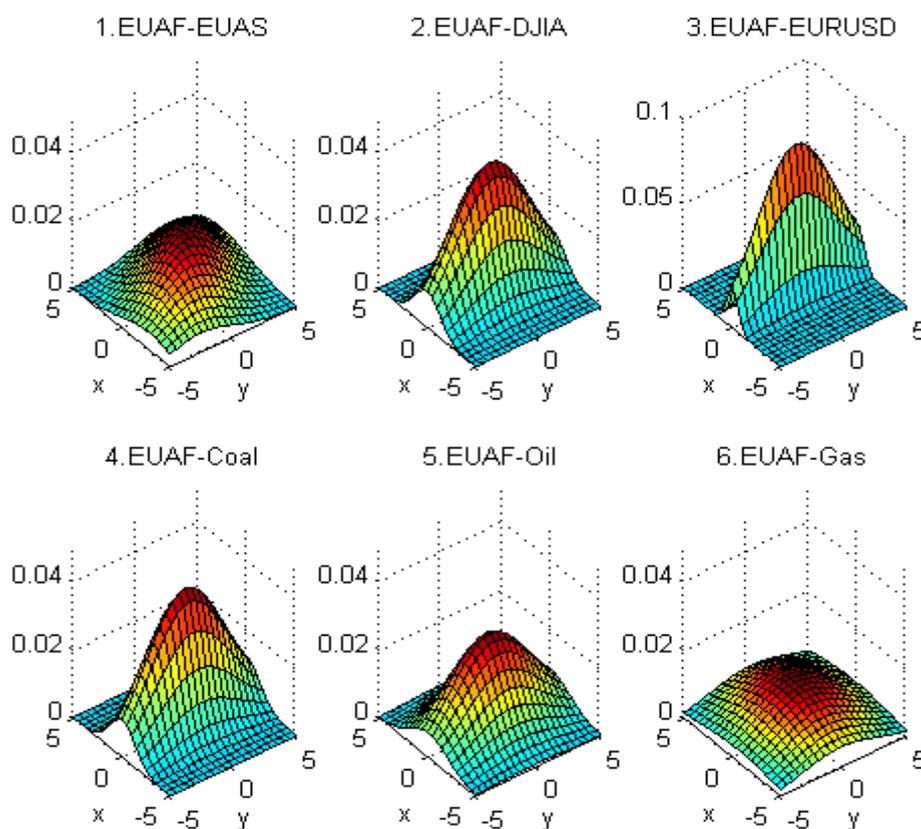


Figure 2. The joint probability density of carbon market and its related markets.

4.2. Designing the Structure of Multi-LSTM

The number of hidden layers is a critical parameter for designing the Multi-LSTM network. However, there is no mature method to determine the number of hidden layers for a deep network, although extensive research by Le and Bengio [28] suggests that several hidden layers are better than one. Correspondingly, this paper applied an experimental method to determine the number of hidden layers.

In this experiment, six levels of the hidden layer number ranging from 1 to 6 and hidden nodes 4, 8, 16, 32, 64, and 128 were used according to the research experience of Shen et al. [29]. In carbon price forecasting, the number of input nodes corresponds to the number of past observations, that is, the number of carbon pricing factors return series and their co-skewness and co-kurtosis terms. Meanwhile, the number of output nodes is set to 1, because the carbon return series is the prediction target. The sample data were divided into two parts in our experiment, that is, 80% of samples were used to train model parameters, and the other 20% were used to test the performance of the model in predicting carbon returns. The empirical results, as shown in Table 2, suggest that the Multi-LSTM model with two hidden layers has the lowest RMSE and MAE values, which can effectively reflect the carbon pricing model under the shock of higher-order moment.

Table 2. Effects of pricing factors on the performance of the Multi-LSTM: Hidden layers and hidden nodes.

Hidden-Layer	Nodes	Multi-LSTM (Under no Shock)		Multi-LSTM (Under the Shock of Higher-Order Moment)		Hidden-Layer	Nodes	Multi-LSTM (Under no Shock)		Multi-LSTM (Under the Shock of Higher-Order Moment)	
		RMSE	MAE	RMSE	MAE			RMSE	MAE	RMSE	MAE
1	4	0.15940	2.34308	0.12987	1.75720	4	4	0.16562	2.39067	0.16668	2.23491
1	8	0.18523	2.68753	0.17476	2.68161	4	8	0.18045	2.64459	0.21730	2.83494
1	16	0.19722	3.01226	0.20264	2.98733	4	16	0.18083	2.76479	0.18154	2.70020
1	32	0.18242	2.77370	0.14342	2.14856	4	32	0.16285	2.52626	0.12551	1.87687
1	64	0.19030	2.93094	0.16401	2.42667	4	64	0.15792	2.39468	0.11356	1.61738
1	128	0.18116	2.82195	0.14156	2.05798	4	128	0.14833	2.23601	0.10978	1.56807
	Avg	0.18262	2.76158	0.15938	2.34323		Avg	0.16600	2.49283	0.15240	2.13873
2	4	0.16354	2.36104	0.13898	1.86684	5	4	0.17004	2.41943	0.17653	2.42557
2	8	0.18178	2.65126	0.16117	2.41483	5	8	0.18223	2.73835	0.15457	2.28414
2	16	0.18136	2.73410	0.16164	2.19602	5	16	0.18396	2.77295	0.14288	2.01281
2	32	0.17152	2.67279	0.12308	1.81584	5	32	0.15333	2.37636	0.12543	1.86661
2	64	0.15452	2.27138	0.10750	1.55172	5	64	0.15252	2.27300	0.10579	1.54763
2	128	0.13723	2.15057	0.11375	1.61574	5	128	0.14845	2.28743	0.11277	1.65257
	Avg	0.16499	2.47352	0.13435	1.91017		Avg	0.16509	2.47792	0.13633	1.96489
3	4	0.16568	2.44213	0.15665	2.18888	6	4	0.16489	2.35490	0.15567	2.40574
3	8	0.16294	2.48370	0.16590	2.39876	6	8	0.17325	2.61837	0.16984	2.43092
3	16	0.18859	2.81335	0.15031	2.27595	6	16	0.18692	2.69322	0.14451	2.15412
3	32	0.18101	2.69656	0.12137	1.78022	6	32	0.16343	2.51814	0.11831	1.66406
3	64	0.15018	2.32166	0.11092	1.60674	6	64	0.15512	2.35640	0.12510	1.79310
3	128	0.15479	2.38795	0.11470	1.67713	6	128	0.14921	2.24498	0.10954	1.57611
	Avg	0.16720	2.52423	0.13664	1.98795		Avg	0.16547	2.46433	0.13716	2.00401

Note: Bold numbers are the minimum root-mean-square error (RMSE) and mean absolute percentage error (MAPE), respectively. MAE, mean absolute error.

The number of hidden layers in the Multi-LSTM network reflects the non-linear processing ability to fit the input data. By using the LSTM network with a few hidden layers, it may be difficult to fit the input data. Meanwhile, it may also be unreasonable to design too many hidden layers, as this may waste a lot of training time and may easily lead to the model being overfitted. The results shown in Table 2 report that the best performance regarding RMSE and MAPE occurs with two hidden layers, because the values of RMSE and MAE are 0.13435 and 1.91017, respectively, which is the lowest value of the entire alternate hidden layer. Further analysis shows that there exists the smallest RMSE and MAE when the number of hidden layer nodes is 64, with the values of 0.1075 and 1.55172, respectively. Therefore, the optimal number of hidden nodes in the Multi-LSTM under the shock of higher-order moment is 64, and the appropriate architecture of Multi-LSTM is 18–64–64–1, where 18 is the number of input nodes and 1 is the output node.

For comparison, we also trained the network structure of the pricing model without considering the shock of higher-order moment, and the optimal architecture is 18–128–128–1, which is different from the structure of Multi-LSTM under the shock of higher-order moment.

4.3. Discussing the Performance of the Proposed Pricing Model

For evaluating the performance of Multi-LSTM in forecasting the carbon price under the shock of higher-order moment of pricing factors, in this section, we take the following classifier as the benchmark model for comparison.

The first is the BP (Back Propagation) model, which is a kind of neural network structure with error backpropagation algorithm and has the ability of non-linear mapping. We took it as a baseline to fit the non-linear relationship of the carbon price and its pricing factors. The second is the GARCH model, which has an advantage in simulating the volatility of the carbon price series. This model has been proven to perform well in previous research. The third one is the Multilayer Perceptron model (MLP), which is a deep neural network structure with many hidden layers, and it can improve the non-linear mapping ability and prediction accuracy in predicting the carbon price. The fourth one is the recurrent neural network (RNN), which is a kind of neural network with special memory ability, and has a remarkable advantage in processing time series data. The fifth is the gated recurrent unit (GRU), which belongs to another improved structure of the RNN model compared with the Multi-LSTM. The GRU model combines the forget gate and the input gate into a single update gate compared to the LSTM structure.

During the experiment, we tested the benchmark models of BP, GARCH, MLP, RNN, and Multi-GRU in forecasting the carbon price with the same series data. That is, we studied the out-of-sample performance of the proposed model and the other comparative models in carbon price prediction. During the experiment, the comparison models maintained the same number of input nodes, hidden layers, hidden nodes, and output nodes as the Multi-LSTM model. The actual return and forecasting return of the different models are shown, as in Figures 3 and 4.

The results presented in Table 3 show that the Multi-LSTM model exhibits better accuracy than the other benchmark models in the five criteria for predicting carbon price when considering the shock of higher-order moment. Specifically, RMSE, MAE, and MAPE represent the stability of the model. The values of the three variables (that is, 1.7347, 0.9884, and 0.0575, respectively) of the Multi-LSTM model are significantly lower than those of the other benchmark models. Therefore, we can conclude that the Multi-LSTM model is more superior in stability and reliability for forecasting the carbon price. As for DA, the direction accuracy of the Multi-LSTM model is 0.9189, that is, significantly higher than other models, and the results prove that the Multi-LSTM model presents more accurate prediction for supporting investor decisions. The correlation coefficient for the Multi-LSTM model is 0.954, which is better than that of Multi-GRU (0.9083), RNN (−0.068), MLP (0.0357), GARCH (0.034), and BP (−0.0332).

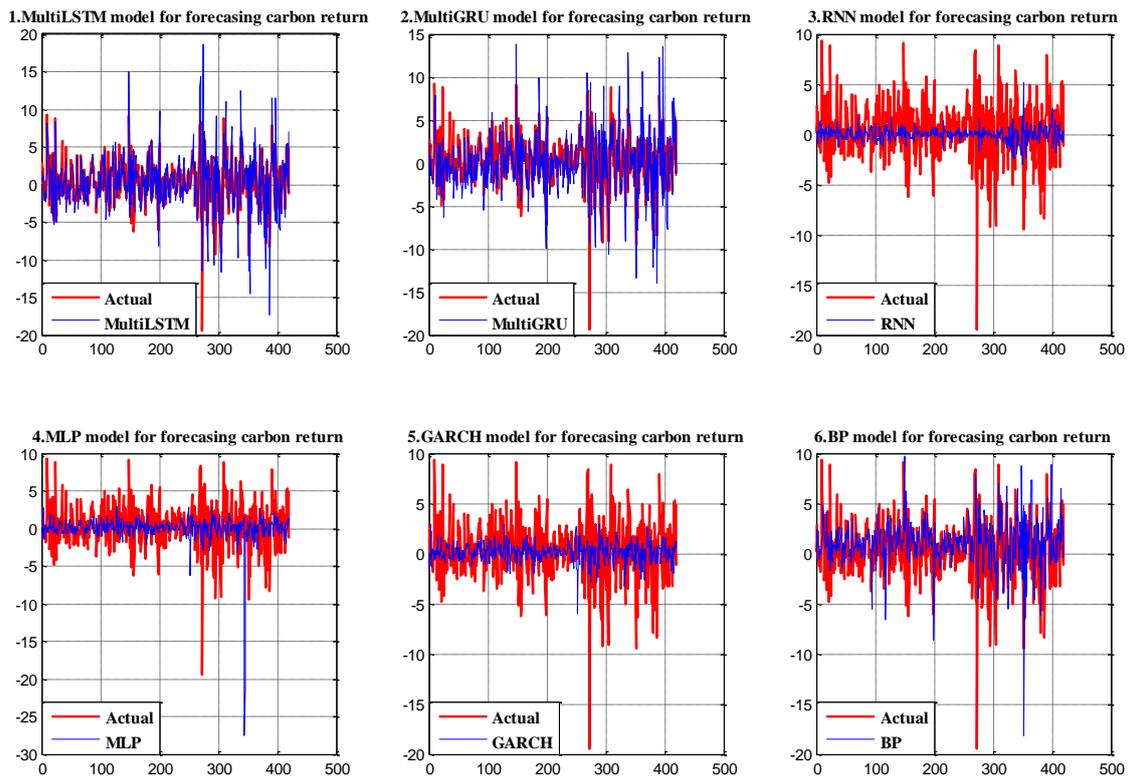


Figure 3. The actual value and prediction value of the proposed model and benchmark model under the shock of higher-order moment. GRU, gated recurrent unit; RNN, recurrent neural network; MLP, multilayer perceptron model; GARCH, generalized autoregressive conditional heteroskedasticity model; BP, backpropagation.

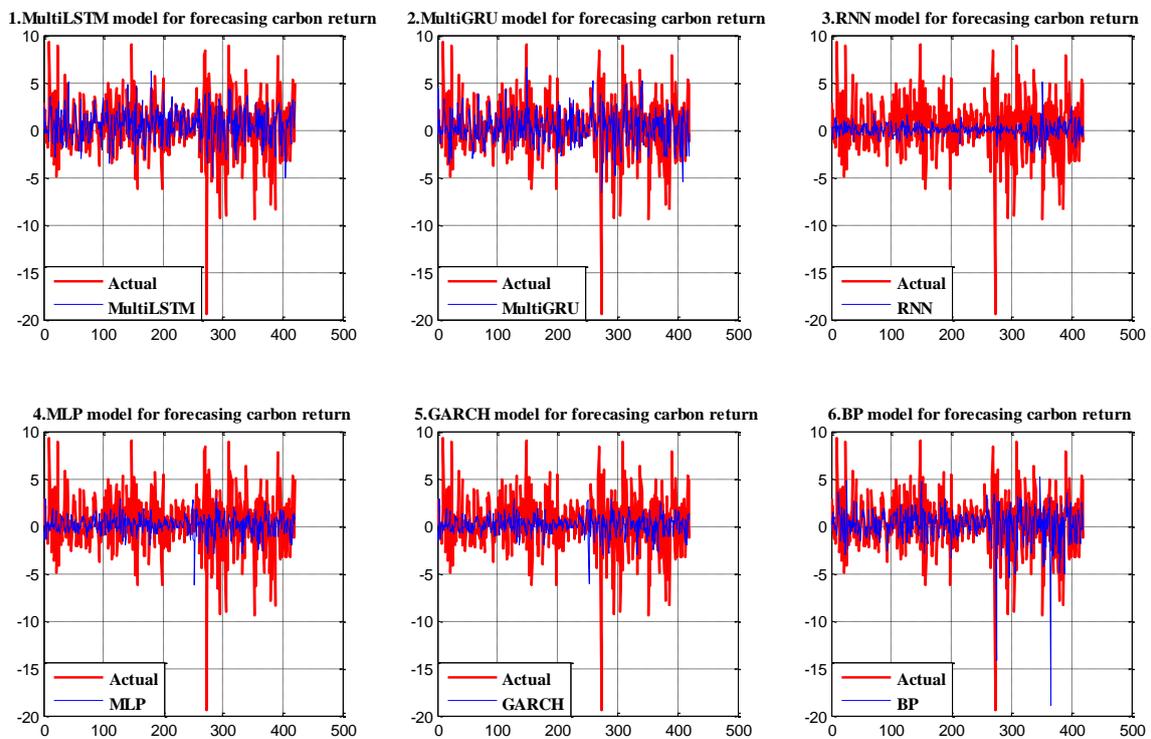


Figure 4. The actual value and the prediction value of the proposed model and the benchmark models without considering the shock of higher-order moment.

Table 3. The performance of the proposed model and the benchmark models for predicting the carbon price.

	Performance Criteria	The Proposed Model		The Benchmark Model			
		Multi-LSTM	Multi-GRU	RNN	MLP	GARCH	BP
Considering the shock of higher-order moment	RMSE	1.7347	1.8918	3.1962	3.6153	3.1783	4.0036
	MAE	0.9884	1.2682	2.3548	2.4955	2.3833	2.9328
	MAPE	0.0575	0.0632	0.1923	0.1246	0.1365	0.0732
	DA	0.9189	0.8711	0.4749	0.5346	0.5442	0.3914
	CORR	0.9540	0.9083	−0.068	0.0357	0.034	−0.0332
Not considering the shock of higher-order moment	RMSE	3.0903	3.1736	3.1950	3.1771	3.1775	3.7228
	MAE	2.2847	2.3517	2.3534	2.3827	2.3826	2.7083
	MAPE	0.0208	0.2731	0.1893	0.1346	0.1355	0.2138
	DA	0.5429	0.5405	0.4714	0.4762	0.4757	0.4357
	CORR	0.2446	0.1926	−0.0847	0.0206	0.0191	−0.0785

Note: DA (direction accuracy) means the probability that the market trend is consistent with the correctly predicted direction. CORR represents the Pearson correlation coefficient between the true carbon return and the predicted return.

The performance of the carbon pricing model without considering the shock of higher-order moment is also listed for comparison. As shown in Table 3, the superiority of the Multi-LSTM model is further proven in all of the five criteria. Figure 5 reports the prediction error curve of the proposed model and the benchmark models when both considering and not considering the shock of higher-order moment, respectively. It is noteworthy that the values of RMSE, MAE, and MAPE for the Multi-LSTM model when considering the shock of higher-order moment are significantly lower than those when not considering the shock of higher-order moment. Meanwhile, the values of DA and CORR for the Multi-LSTM model under the shock of higher-order moment are significantly larger than those when not considering the shock of higher-order moment. Therefore, we can draw a unified conclusion that the Multi-LSTM model based on the proposed EHM-APT framework can significantly improve the prediction accuracy when considering the shock of higher-order moment compared with the pricing models that do not consider the shock of higher-order moment.

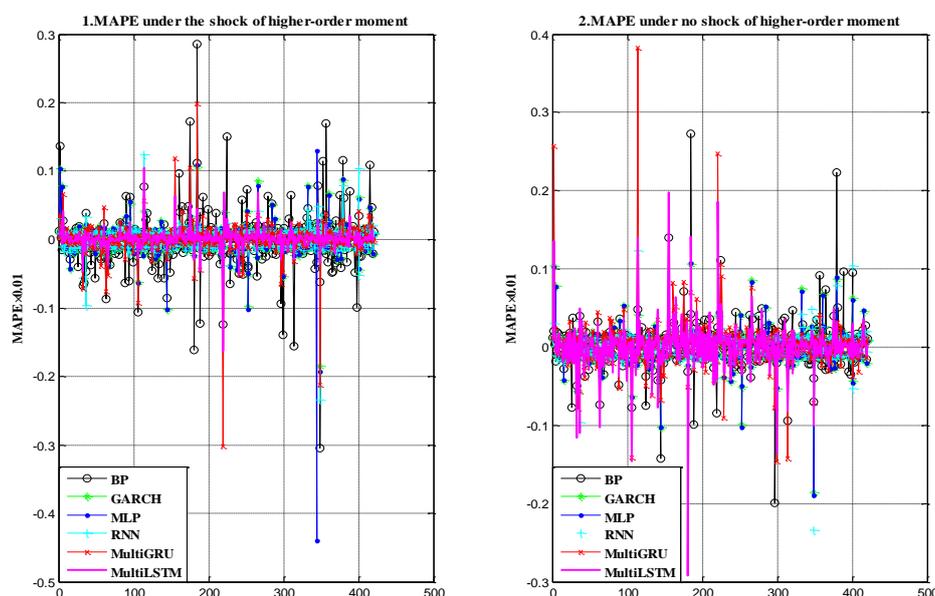


Figure 5. Prediction error curve of the proposed model and the benchmark models. Note: The horizontal axis represents the sample sequence, and the vertical axis uses RMSE to represent the prediction error of the Multi-LSTM and benchmark models. Figure 5 shows the performance of the different models visually, and highlights the superiority of multi-LSTM in fitting the proposed EHM-APT framework.

5. Conclusions and Prospects

In this paper, we developed an extended higher-order moment Multi-factor framework (EHM-APT) for predicting the carbon price by extending the theory of the binary higher-order moment CAPM model developed by Fry et al. [6] to multivariate factors. The proposed EHM-APT framework in this paper not only helps to promote the construction of an effective carbon pricing mechanism, but also provides a more effective market-oriented means for carbon emission reduction and the sustainability of low-carbon economy. As for the EHM-APT framework, we considered the impact of market asymmetry and extreme events that stem from its pricing factor markets on the carbon price compared with the traditional APT model. Furthermore, a multi-layer and multi-variable LSTM (Multi-LSTM) model was constructed so that the parameters and structure can be determined experimentally for testing the performance of the proposed EHM-APT framework in predicting the carbon price. Some main conclusions are summarized as follows.

Firstly, the proposed EHM-APT framework, when considering the shock of higher-order moment, can significantly improve the prediction accuracy of the carbon price compared with the framework when not considering the shock higher-order moment. This conclusion further proves that the shock of market asymmetry and extreme events that stem from its pricing factors is an indispensable factor for predicting and fitting the carbon price, which is ignored by previous research. In fact, it has been proven in many studies that co-skewness and co-kurtosis, which represent the market asymmetry and extreme events, are used as the pricing factors of financial assets [4]. Based on this idea, this paper further proves that co-skewness and co-kurtosis are also important and indispensable factors for explaining the pricing mechanism of carbon assets.

Secondly, the Multi-LSTM model for which the parameters and structure were determined experimentally (the structure of 18–64–64–1) in this article, is a suitable network for predicting the carbon price, and its performance is superior to the other benchmark models (that is, the Multi-GRU, RNN, MLP, GARCH, and BP models) in all of the evaluation criteria. This result suggests that the Multi-LSTM model is competitive for predicting the carbon price under the proposed EHM-APT framework. Therefore, we conclude that the model has advantages in effectively capturing and mapping the complex non-linear network relationship between the carbon price and its pricing factors.

This study predicts the carbon price based on the structured pricing factors that the data sourced from certain financial markets. As evidenced in this article, the proposed EHM-APT framework has good applicability, and can support valuable references for solving pricing problems in other financial markets as well. In fact, in addition to the higher-order moment terms, we can also explore pricing factors that can measure market asymmetry information and external shock by means of big data text mining. Future research can explore the non-structured pricing factors obtained by text information technology by applying the proposed multi-factor pricing framework, such as the factor of investor sentiment, policy events, and others. Based on this, another valuable exploration prospect is derived, that is, as the increase of data information by means of text mining technology or other machine learning technologies, the way to achieve information fusion and sharing of data features between the structured and unstructured pricing factors within the established pricing framework is a key issue that requires an outcome.

Author Contributions: Conceptualization, P.Y. and C.Z.; Methodology, P.Y. and Y.W.; Model code analysis, P.Y. and X.Y.; Formal Analysis, Y.W. and X.Y.; Writing—Original Draft Preparation, P.Y. and Z.A.W.; Writing—Review and Editing, P.Y. and C.Z.; Supervision, C.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This paper is supported by the following fund projects: the National Natural Science Foundation of China, No. 71971071 and 71373065.

Acknowledgments: The authors thank the editors and reviewers for their attention and concern.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Schneider, L.; La Hoz Theuer, S. Environmental integrity of international carbon market mechanisms under the Paris Agreement. *Clim. Policy* **2019**, *19*, 386–400. [[CrossRef](#)]
2. Yang, B.; Liu, C.; Gou, Z.; Ma, J.; Su, Y. How Will Policies of China's CO₂ ETS Affect its Carbon Price: Evidence from Chinese Pilot Regions. *Sustainability* **2018**, *10*, 605. [[CrossRef](#)]
3. Hammoudeh, S.; Lahiani, A.; Nguyen, D.K.; Sousa, R.M. An empirical analysis of energy cost pass-through to CO₂ emission prices. *Energy Econ.* **2015**, *49*, 149–156. [[CrossRef](#)]
4. Amaya, D.; Christoffersen, P.; Jacobs, K.; Vasquez, A. Does realized skewness predict the cross-section of equity returns. *J. Financ. Econom.* **2015**, *118*, 135–167. [[CrossRef](#)]
5. De Luca, G.; Loperfido, N. Modelling multivariate skewness in financial returns: A SGARCH approach. *Europ. J. Financ.* **2015**, *21*, 1113–1131. [[CrossRef](#)]
6. Fry, R.; Hsiao, C.Y.L. Extremal dependence tests for contagion. *Econom. Rev.* **2016**, *37*, 626–649.
7. Karmakar, M.; Paul, S. Intraday risk management in international stock markets: A conditional EVT approach. *Int. Rev. Financ. Anal.* **2016**, *44*, 34–55. [[CrossRef](#)]
8. Chevallier, J. Carbon futures and macroeconomic risk factors: A view from the EU ETS. *Energy Econ.* **2009**, *31*, 614–625. [[CrossRef](#)]
9. Byun, S.J.; Cho, H. Forecasting carbon futures volatility using GARCH models with energy volatilities. *Energy Econ.* **2013**, *40*, 207–221. [[CrossRef](#)]
10. Koop, G.; Tole, L. Forecasting the European carbon market. *J. R. Stat. Soc. A Stat.* **2013**, *176*, 723–741. [[CrossRef](#)]
11. Sanin, M.E.; Violante, F.; Mansanet-Bataller, M. Understanding volatility dynamics in the EU-ETS market. *Energy Policy* **2015**, *82*, 321–331. [[CrossRef](#)]
12. Benz, E.; Stefan, T. Modeling the price dynamics of CO₂ emission allowances. *Energy Econ.* **2009**, *31*, 4–15. [[CrossRef](#)]
13. Chevallier, J. Nonparametric modeling of carbon prices. *Energy Econ.* **2011**, *33*, 1267–1282. [[CrossRef](#)]
14. Zhang, Y.J.; Sun, Y.F. The dynamic volatility spillover between European carbon trading market and fossil energy market. *J. Clean. Prod.* **2016**, *112*, 2654–2663. [[CrossRef](#)]
15. Wu, Q.; Wang, M.; Tian, L. The market-linkage of the volatility spillover between traditional energy price and carbon price on the realization of carbon value of emission reduction behavior. *J. Clean. Prod.* **2020**, *245*, 1–12. [[CrossRef](#)]
16. Ji, Q.; Zhang, D.; Geng, J.B. Information linkage, dynamic spillovers in prices and volatility between the carbon and energy markets. *J. Clean. Prod.* **2018**, *198*, 972–978. [[CrossRef](#)]
17. Oliveira, F.A.; Nobre, C.N.; Luis, E.Z. Applying artificial neural networks to prediction of stock price and improvement of the directional prediction index-Case study of PETR4, Petrobras, Brazil. *Expert Syst. Appl.* **2013**, *40*, 7596–7606. [[CrossRef](#)]
18. Garcia, F.; Guijarro, F.; Oliver, J.; Tamosiuniene, R. Hybrid Fuzzy Neural Network to Predict Price Direction in the German Dax-30 Index. *Technol. Econ. Devel. Econom.* **2018**, *24*, 2161–2178. [[CrossRef](#)]
19. Zhu, B.Z.; Wei, Y.M. Carbon price prediction based on integration of GMDH, particle swarm optimization and least squares support vector machines. *Syst. Eng. Theory Pract.* **2011**, *31*, 2264–2271.
20. Fan, X.; Li, S.; Tian, L. Chaotic characteristic identification for carbon price and an multi-layer perceptron network prediction model. *Expert Syst. Appl.* **2015**, *42*, 3945–3952. [[CrossRef](#)]
21. Zhu, B.; Ye, S.; Wang, P.; He, K.; Zhang, T.; Wei, Y.M. A novel multiscale nonlinear ensemble leaning paradigm for carbon price forecasting. *Energy Econ.* **2018**, *70*, 143–157. [[CrossRef](#)]
22. Atsalakis, G.S. Using computational intelligence to forecast carbon prices. *Appl. Soft. Comput.* **2016**, *43*, 107–116. [[CrossRef](#)]
23. Zhu, J.; Wu, P.; Chen, H.; Liu, J.; Zhou, L. Carbon price forecasting with variational mode decomposition and optimal combined model. *Phys. A* **2019**, *519*, 140–158. [[CrossRef](#)]
24. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
25. Qiu, M.; Song, Y. Predicting the direction of stock market index movement using an optimized artificial neural network model. *PLoS ONE* **2016**. [[CrossRef](#)]
26. Moghaddam, A.H.; Moghaddam, M.H.; Esfandyari, M. Stock market index prediction using artificial neural network. *J. Econom. Financ. Admin. Sci.* **2016**, *21*, 89–93. [[CrossRef](#)]

27. Hwang, S.; Satchell, S.E. Modelling emerging market risk premia using higher moments. *Int. J. Financ. Econom.* **1999**, *4*, 271–296. [[CrossRef](#)]
28. Le, R.N.; Bengio, Y. Representational power of Restricted Boltzmann Machines and Deep Belief Networks. *Neural Comput.* **2008**, *20*, 1631–1649.
29. Shen, F.; Chao, J.; Zhao, J. Forecasting exchange rate using deep belief networks and conjugate gradient method. *Neurocomputing* **2015**, *167*, 243–253. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).