Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy

Pol. J. Environ. Stud. Vol. 30, No. 5 (2021), 1-14
DOI: *10.15244/pjoes/131856* ONLINE PUBLICATION DATE:

Original Research

Forecasting China's Steam Coal Prices Using Dynamic Factors and Mixed-Frequency Data

Chunyang Wang[1], Wanglin Kang[2]*

[1]School of Business, Macau University of Science and Technology, China
[2]School of Economics and Management, Shandong University of Science and Technology, Qingdao, China

Received: 12 October 2020
Accepted: 21 December 2020

Abstract

This paper investigates the dynamic relationship between steam coal price and its drivers sampling mixed frequencies to improve the prediction of weekly steam coal price. A novel hybrid method, combining the mixed data sampling (MIDAS) model with eXtreme Gradient Boosting (XGBoost) algorithm, is proposed to perform forecast of weekly steam coal prices by applying the latest mixed factors with high frequencies. The empirical evidences indicate that the daily natural gas prices, temperatures, and air quality index (AQI) have better predictive abilities for steam coal prices than the A-share index and crude oil prices. It's shown that the hybrid model has approximately 23.27% and 78.39% accuracy improvement over the combination-MIDAS and other benchmark models, respectively. The empirical results are helpful for the government to effectively capture the fluctuation and uncertainty of steam coal prices from the energy market and environmental conditions to make reasonable strategies in China.

Keywords: steam coal price, MIDAS model, forecast combination, XGBoost algorithm

Introduction

As the world's largest energy consumer, China has consumed 4.64 billion tons of coal equivalent with coal accounting for 59% of its energy mix in 2018. More than 70% of coal is used to generate electricity production in recent years [1]. Except for the human resource and fix input, steam coal prices are an essential part of the cost of coal-based power plants. In order to improve the efficiency of electricity generation and steel manufacturing in coal consumption [2], China has established the national coal trading market, i.e., Qinhuangdao coal trading market in 1992. With the reform deepening, Qinhuangdao coal price systems including steam coal price were issued in 2007 [3]. However, the immaturity of the steam coal market of China and the impact of international trade uncertainty make the steam coal price more volatile. Hence, it is urgent to present an effective forecasting technology for steam coal prices, which is the foundation to provide the price reference for the government's macro-regulations and tool for power enterprise's risk aversion.

The current literature has explored the key factors that affect steam coal prices from coal production,

*e-mail: wlkang@163.com

Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy

energy market, and economic development. The costs of coal production involving the coal resource supply, investment, technology, and transportation are now a substantial influence on steam coal prices [4]. From the theory of supply and demand, Ding et al. pointed out that China's future coal supply and demand trend is increasing, and correspondingly, coal prices will drop first and then rise [5]. From the energy side, some studies investigate that the price trends of the oil, natural gas, and steam coal price [6-7] are the same in the long term [8]. All over the world, oil as the most important energy source has become the benchmark by which other energy sources are measured. The price of coal, as another crucial fossil fuel, has been historically strongly related to oil. The greenhouse climate change motivates the national governments to use clean energies to replace the coal. Natural gas is considered as clean and cost-effective alternatives for electricity generation or coal-fired power plants [9-10]. The evidence shows that there exists the inter-market contagion of the international energy market, where the steam coal prices are influenced by fossil energy, e.g., oil, coal, and natural gas markets [11]. Some literatures also point out that there is a significant bi-directional volatility spillover between the clean energy stocks and the steam coal market [12]. On the contrary, Liu et al. believed that deregulation of power industry will lead to lower coal prices [13].

Economic development is also considered as one of the dominating factors in the determination of steam coal prices. In developing countries, such as China and India, are experiencing the rapid economic development of their industries, with 6.1% and 5.3% GDP growth, respectively. India has become one of the growing countries next to China. To meet the need for this rapid growth, the generation of power is needed at an exorbitant rate in India [14], and 75% of its electricity is currently generated by coal-fired power plants [15]. Similarly, some evidence also shows that economic growth stimulates coal consumption in China [16]. With the economic development, the rapidly increased demand for electricity in China has boosted the coal-fired power plant capacity. Hence, along with its economic development, China has sparked growth among power, steel, and chemical industries relying on coal as a feedstock, which could lead to the fluctuation of the steam coal price.

Although the current literature has realized the importance of managing steam coal price and identified the influencing factors, the limited research pays attention to steam coal price prediction. For example, Zhan and Ma used the modified partial least-squares method (LSM) to forecast the daily coal price indexes in Qinhuangdao of China [17]. Krzemień et al. presented the generalized regression neural network (GRNN) to forecast the European steam coal spot price [18]. Zhao et al. examined the fluctuations of steam coal price in China by using the multi-fractal detrended fluctuation analysis (MFDFA). They pointed out that the quarterly fluctuation index (QFI) had a better forecasting ability when the prices fluctuated wildly [19]. Fan et al. established a multi-layer perceptron network model (MPNM) to make short terms predictions for the coal prices of Qinhuangdao of China [20]. Alameer et al.combined the long short-term memory (LSTM) with deep neural network (DNN) to forecast monthly coal price fluctuations in China [21]. Since the coal consumptions directly affect the coal prices, the coal consumption prediction also achieves attention. Wang et al. combined the grey model (GM) with the autoregressive integrated moving average model (ARIMA) to predict the coal consumption in U.S.A [22]. Wang et al. established a hybrid model based on particle swarm optimization (PSO) to forecast the coal demand in China [23].

The above literature has developed various models to coal price prediction, including linear methods, e.g., ARIMA and nonlinear methods, e.g., LSTM [14]. In order to solve some disadvantages in single forecasting method, the combined models are constructed, such as LSTM- DNN and GM- ARIMA [21-22].

But these forecasting methods and models only consider the daily coal price forecast. In fact, the daily volatility of spot coal prices sometimes was very low, or even unchanged. For example, in the European DES ARA (Delivered Ex-ship; Amsterdam, Rotterdam, Antwerp) market, the weekly or monthly coal price data are often released. Hence, it needs to explore the forecast technology to predict the low-frequency coal prices using daily driving factors. Traditionally, these daily driving factors are always converted into lower frequency data, e.g., weekly data by averaging or bridging methods. This may result in the loss of valuable information from the different frequency data and inaccuracy of forecast. In recent years, the mixed data sampling (MIDAS) approach, proposed by Ghysels et al. [24] allows the explanatory variables and the dependent variables to be sampled at different frequencies, while distributed lag polynomial is used to ensure parsimonious specifications [25]. This method is proven to be an effective way to perform real-time forecasts for dynamic prices by employing the more abundant sample information at different frequencies [26-27].

However, there still exist some research gaps in the prediction of steam coal prices by using key factors at different frequencies. There is no attempt to apply the MIDAS regression model to forecast the steam coal prices. Some scholars have criticized that the single use of MIDAS models can not achieve optimal performance in all conditions because of sampling variation, structural breaks, and random factors [28]. Hence, this paper proposes a hybrid model to forecast the steam coal prices. In order to get a better grasp of the dynamic and nonlinear features of weekly steam coal prices, we combine the eXtreme Gradient Boosting (XGBoost) method with MIDAS models to improve the accuracy. The technique of XGBoost has been proved

Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy

to have a better capability of extracting meaningful information from unstructured data [29]. The evidence shows that the XGBoost method has an advantage over the random forest, Bayesian, and k-Nearest Neighbors models in terms of speed and prediction accuracy [30]. The unique respective strengths of MIDAS model and XGBoost method contribute to present the flexible non-linear forecasting capabilities for steam coal prices.

This paper aims to investigate the dynamic relationship between weekly steam coal prices and daily drivers to improve the accuracy of steam coal price prediction. The contributions are presented as follows. First, the comprehensive factor systems are presented not only including the daily crude oil prices, the natural gas prices, and the A-share index, but also the daily temperature and air quality index (AQI). Second, this paper constructs a novel hybrid model, combing the MIDAS model with the XGBoost algorithm to perform a forecast of weekly steam coal prices by applying the latest mixed factors with higher frequencies. The hybrid model can predict the errors to capture the nonlinear changes of weekly steam coal prices, which helps to correct the forecast of steam coal prices from the best individual model. The empirical findings indicate that natural gas prices, temperatures and AQI have better predictive abilities than the A-share index and crude oil prices. Third, compared with the benchmark models, e.g., autoregression (AR), moving average (MA), and ARMA, the novel hybrid model achieves better performance. It's shown that the hybrid model has approximately 23.27% and 78.39% accuracy improvement over the combination-MIDAS and benchmark models, respectively. The results are beneficial to the policymakers' participants in environment protection and energy investment decisions.

The remaining part of the paper is organized as follows. Section 2 introduces the hybrid model in this paper. Section 3 describes the data resources and processing. The empirical results are presented in section 4 and section 5. Section 6 reports the research conclusions and future work.

Material and Methods

The MIDAS method has wide applications for multiple forecasting domains, such as prediction of the financial market [31-32], energy market [33-34], and other macroeconomic issues, e.g., GDP [35-36] The MIDAS method can fully utilize high-frequency data without sustainable loss of sample information to directly reflect the dynamic relationships among variables by polynomial weights [37]. To realize the accurate forecast for steam coal price, and to address the misspecification of the individual MIDAS model [38], this paper constructs the combination-MIDAS regression model with the comprehensive driving factor systems.

The MIDAS Model

The existing literature discussed early shows that steam coal prices can be influenced by many factors, such as energy markets, economic markets, weather, and environmental aspects. Since the individual MIDAS model can not use multiple indicators, the combination-MIDAS regression model was proposed to make forecasts under different information sets and indicators [39]. For example, the combination-MIDAS model has been used to forecast crude oil price [40], carbon price [41], and wastewater discharge [42].

The critical step to construct a combination-MIDAS model is to present the formation of a combined forecast based on a series of individual forecasts [38]. It is common to use weight techniques to create the formation of a combined forecast by employing the information sets of the individual models [43]. Given N forecasting results of multiple individual MIDAS models, the final forecast result of the combination-MIDAS model is defined as follows.

$$\overline{f}_{N,T+h|T} = \sum_{j=1}^{N} (w_{j,T} \times \overline{y}_{j,T+h|T}) \quad (1)$$

...where T refers to the last observation in the estimator sample for the best individual MIDAS, $w_{j,T}$ refers to the combination weights formed at time T, and $\overline{y}_{j,T+h|T}$ refers to jth individual forecast result of $\overline{y}_{j,T+h|T}$ computed by the best individual model in the out-of-sample at time T. $\overline{f}_{N,T+h|T}$ denotes the forecast combination generated at time T, which is a weighted average of N individual h-step ahead forecast results in the out-of-sample. In this paper, the combination-MIDAS model is formed with five weight techniques, e.g., MSFE-weighted type, DMSFE-weighted type, AIC-weighted type, BIC-weighted type, Equal-weighted type. For details of these five weight techniques, please see Appendix A.1. To realize forecasting with the latest available published data and to consider the autoregressive effect of Y_t, the ADL-MIDAL (m, k, h) model with h-step-ahead is constructed as follows.

$$Y_t = \alpha + \sum_{j=1}^{p} \gamma_j Y_{t-j} + \beta W(L^{1/m}, \theta) X_{t-h/m}^{(m)} + \varepsilon_t \quad (2)$$

...where Y_t refers to the weekly steam coal prices and t = 1, 2, ···, T. $X_t^{(m)}$ refers to the ith daily indicators, which can be observed m times between week t – 1 and t. Thus, set m = 5 in this paper. Here, α, γ_j, β are the model parameters, h refers to the leads of daily factors, and ε_t is the stochastic disturbance term. When h is greater than or equal to 1, this model can use the daily factors before Friday to forecast ahead of the steam coal price. The term of $W(L^{1/m}, \theta)$ consists of the lag operator $L^{1/m}$ and a parameter vector of limited dimension θ. $W(L^{1/m}, \theta)$ can be denoted by $\sum_{k=i}^{K} w(k;\theta) L^{k/m}$, where

Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy

$w(k, \theta)$ is a polynomial weight, and K refers to the maximum lag order of the daily high-frequency variable. Set $L^{k/m}X_t^{(m)} = X_{t-k/m}^{(m)}$ where $L^{k/m}$ is a lag operator. When k = 0, $X_{t-k/m}^{(m)}$ refers to the observation of Friday in a week. When k = 1, $X_{t-k/m}^{(m)}$ refers to the observation of Thursday in a week, and so on.

The MIDAS model depends on the polynomial weights to capture the dynamic relationships among data with different frequencies. Thus, it's important to choose an appropriate weight function for the ADL-MIDAS model to achieve an accurate forecast. To reflect the direct effects of different high-frequency data on the steam coal prices, various parsimonious polynomial specifications of $W(L^{1/m}, \theta)$ have been taken in this paper. These specifications include the beta density function polynomial with zero lag (Beta), the beta density function polynomial with non-zero lag (BetaNN), Exponential Almon lag polynomial (ExpAlmon), Almon lag polynomial (Almon), Step function (Stepfun), and Unrestricted MIDAS (UMIDAS) [39] (see details in Appendix A.2).

XGBoost Algorithm

For the accurate prediction of the steam coal prices, this paper presents the XGBoost algorithm to forecast the change of prediction error [43]. As a scalable machine learning system for tree boosting, XGBoost shows better performance in analyzing and predicting the prices compared with other machine learning models [44], such as support vector machines (SVM) and hybrid associative memory with translation (HACT). Since XGBoost is an optimized distributed gradient boosting algorithm which can capture the nonlinear characters of the data series [45].

XGBoost model could be generated at each iteration, and the residuals can be utilized to modify the previous predictor to generate the optimal loss function. Specifically, the XGBoost model integrates multiple weak classifiers into a strong classifier by a series of specific learning algorithms [46]. Here, the weak classifier is the sub-model produced by a series of the model iteration process. The strong classifier is the final error prediction model where the weak classifiers generated are added by a specific weight function after the completion of iterations.

The prediction error is the final output by summing up the scores in the corresponding decision trees. The function form is
displayed as:

$$\hat{y}_i = \phi(X_i) = \sum_{k=1}^{K} f_k(X_i) \quad f_k \in \Theta \tag{3}$$

...where Θ refers to the space containing the forecasting functions and the output scores of each function, and $f_k(X_i)$ refers to the kth forecasting function in the function space, and $\hat{y}_i$ denotes the final prediction that could capture the short-term nonlinear characters. Here, $X_i = (x_1, x_2, ..., x_m)$ refers to the input variables.

To realize the accurate forecast of the prediction error, we construct the following multi-objective function.

$$\ell(\phi) = \sum_{i=1}^{N} l(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k),$$
$$where \quad \Omega(f_k) = \gamma T + \frac{1}{2}\chi \sum_{t=1}^{T} w_t^2 \tag{4}$$

...where $l(y_i, \hat{y}_i)$ is a differentiable convex loss function referring to the gap between the target error and the prediction error, and $\Omega(f_k)$ is a penalty term added to the loss function to avoid overfitting. Here, γ and χ refers to the penalty factors, T refers to the branch number of each weak classifier, and w_t indicates the weighted score of each branch of the weak classifier. Moreover, the goal of each loss function in this model is to minimize the gap between the prediction error and the target error. $\Omega(f_k)$ measures the complexity of the proposed model where the lower the complexity, the stronger the generalization ability. Therefore, by minimizing the objective function of Equation (4), the best model of predictive function could be selected [46].

Notably, since functions as import parameters enter the target functions Equation (4), it's unreasonable to optimize the prediction model in Euclidean space by the traditional estimate method [43]. Considering the deficiency of the traditional estimate method, second-order Taylor expansion is taken to the loss function of Equation (4). Assumed that the loss function is mean square error (MSE), and thus the multi-objective function in the tth iteration is shown:

$$\ell^{(t)} = \sum_{i=1}^{N} l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t)$$
$$; \sum_{i=1}^{N} [l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t) \tag{5}$$

...where $g_i = \partial_{\hat{y}^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$ and $h_i = \partial^2_{\hat{y}^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$. Here, the additive manner is used to model leaning; that is, at each iteration, the algorithm identifies a new function $f_k(x_i)$ in the functions space Θ. This method can minimize the gap between the target value γ and the prediction value $\hat{y}_i^{(t)}$ at the tth iteration, and thus generate the final loss function.

According to Equation (5), the finally multi-objective function is given as:

$$\ell^* = -\frac{1}{2} \sum_{j=1}^{T} \frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_i + \chi} + \gamma T \tag{6}$$

...where ℓ^* represents the optimal solution of the multi-objective function. Therefore, obtaining the desired

Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy

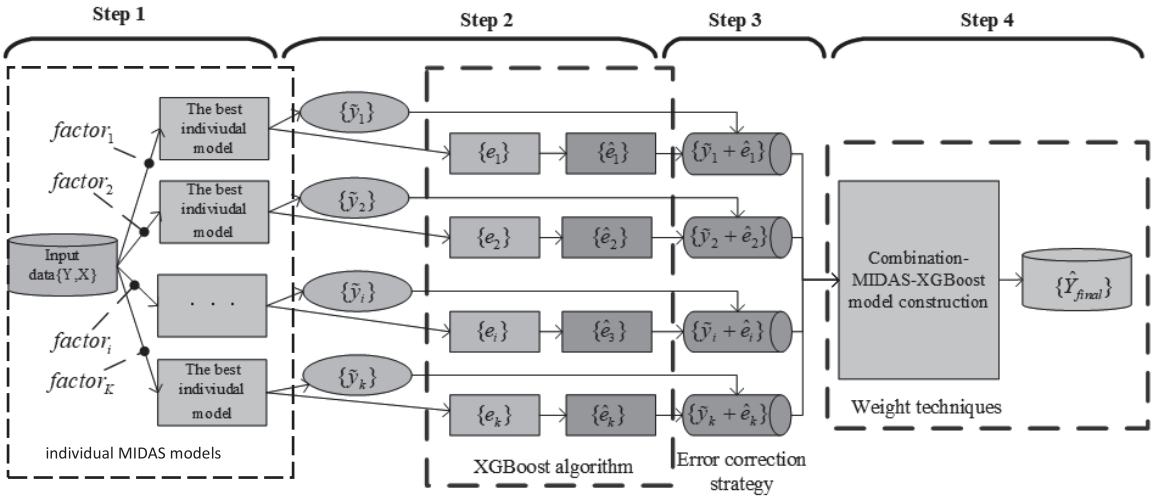

Fig. 1. Flow chart of combination-MIDAS-XGBoost model.

output $\hat{y}_i$ can be transformed into a question that makes the multi-objective function ℓ^* of Equation (6) minimum.

Combination-MIDAS-XGBoost Model

In this paper, the proposed combination-MIDAS-XGBoost model consists of four steps, as shown in Fig. 1. In step 1, the main goal is to establish the individual MIDAS models with each factor and select the best individual model according to its predictive performance in out-of-sample. Here, the root means squared errors (RMSE) are adopted to measure the predictive performance [47]. In step 2, taking the weekly steam coal price, corresponding higher-frequency factors, and the residuals of the individual MIDAS model as the input data, the XGBoost algorithm could be achieved, and the forecasting values of the residuals is the prediction errors. In step 3, the predictive errors generated by the XGBoost algorithm is utilized to correct the forecast values of the individual MIDAS model. The modified forecast performance of the individual model shows higher accuracy. In step 4, the forecast for weekly steam price could be achieved by the combination-MIDAS-XGBoost model.

Data Description

This paper uses the daily driving factors from energy, economy, weather, and environmental aspects to forecast weekly steam coal prices. The data covers the period from November 11, 2013, to November 15, 2019, which is divide into two parts. Namely, 80% of the data is constructed as the training data and 20% of the data is constructed as the out-of-sample data. Specifically, the training data is set to train the proposed hybrid model, of which period from November 11, 2013, to October 5, 2018. The period from October 12, 2018, to November 15, 2019, is utilized for the out-of-sample forecast.

The weekly steam coal price is from China's Qinhuangdao port, which is the largest coal import and export port in the world [20]. The fluctuation of steam coal price in Qinhuangdao port deeply influences the global coal demand and supply, which has an important strategic position in guaranteeing China's energy security (source: Chinese Wind Database).

The daily driving factors from the energy market include the WTI crude oil future price and the NYMEX natural gas future price (source: Chinese Wind Database). The daily economic factor is represented by the A-share index (source: Chinese Wind Database). For weather factors, this paper selects the daily average temperature of China as the proxy (source: National Ocean and Atmospheric Administration). Regarding the environmental factor, this paper selects the daily average air quality index (AQI) in 358 major cities of China (source: Tianqihoubao web of China).

To improve the forecasting accuracy and eliminate heteroscedasticity, the proposed hybrid model adopts the growth rates of variables. Specifically, $growth_{i,t} = ln(value_{i,t}/value_{i,t-1})\times 100$, where $growth_{i,t}$ refers to the return rate or growth rate of the ith variable at time t, and $value_{i,t}$ refers to the ith observation at time t. The return rates or growth rates of variables are plotted in Fig. 2.

Influence of Driving Factors for Steam Coal Price

Selection of the Best Single MIDAS Model

This section selects the best single MIDAS model with a predictor to forecast the steam coal price using the model discussed in section 3.1. The single MIDAS model with the best accuracy and the predictor with

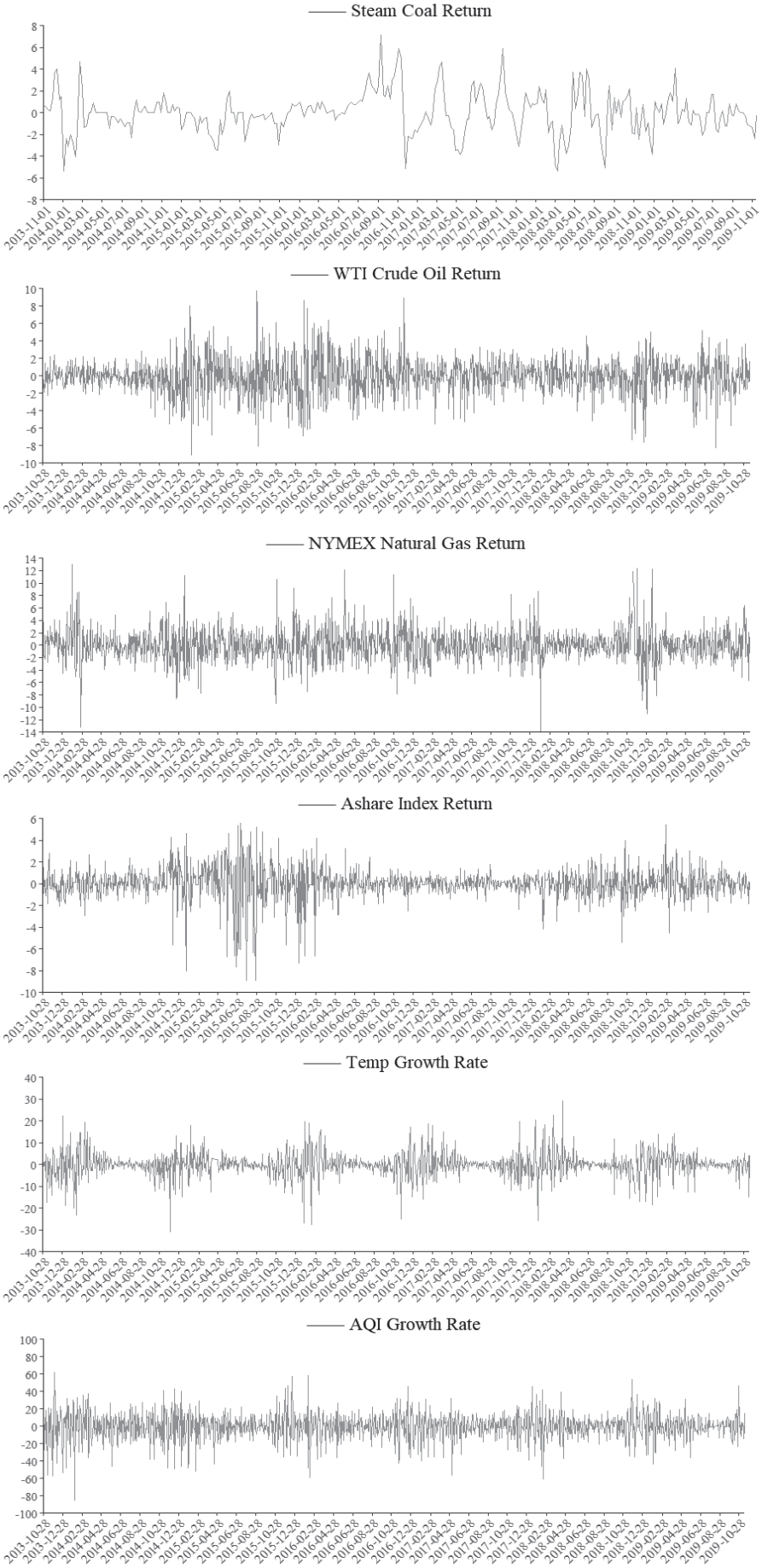

Fig. 2. Return of steam coal price and the growth rates of related factors.

Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy

Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy

the greater predictive ability is selected by comparing the RMSEs value, which is in line with Han et al. [41] As the first step to construct the proposed combination-MIDAS-XGBoost model, that selecting the best MIDAS model is a critical process to the accuracy of the final results because it's the basis of the proposed hybrid model and further influences the final forecasting results. In addition, the excellent predictor also is chosen by comparing the RMSEs value generated by the single MIDAS model.

The best polynomial weight reflecting the dynamic relationship between the steam coal price and the relevant factors, and the best lag order of the corresponding factor are also determined by the RMSEs value generated by the single MIDAS model with fixed window method. To construct the best MIDAS model with accurate prediction and reflect the changing trend in RMSEs value, this paper sets the maximum lag orders of daily factors crude oil, natural gas, A-share, temp, and AQI as 60, set the maximum lag order of the weekly steam coal price as 5 when considering the 0, 1, 2, 3, 4 step ahead conditions. Moreover, several polynomial weights also are considered in this model, which have been shown in Appendix A.2. This paper takes the single MIDAS model with the crude oil price as an instance to make clear the mechanism of selecting the best polynomial weight and the best lag order for the steam coal price and the crude oil price. Table 1 shows the detailed RMSEs value generated by the single MIDAS model with 1 lead of the steam coal price; that is, the RMSEs value of this model when considering h = 1. When 0 lag order for steam coal price included in this model, the polynomial weight of BetaNN performs better in reflecting the relationship between the steam coal price and the crude oil price, and the best lag order for the crude oil price is 54. When the lag order for steam coal price included in this model is 1, 3, respectively, BetaNN performs better in reflecting the relationship between the steam coal price and the crude oil price than other alternative polynomial weights, and the best lag order for the crude oil price is 54. When 2, 4, 5 lag order for steam coal price included in this model,

Table 1. RMSEs generated by the single MIDAS model with 1 lead of the steam coal price.

Weights	Lag orders for crude oil price					Lag orders for crude oil price				
	9	20	33	45	54	9	20	33	45	54
	0 lag orders for steam coal price					1 lag orders for steam coal price				
Beta	1.3808	1.4765	1.4752	1.4746	1.4744	1.4341	1.4341	1.4341	1.4341	1.4341
BetaNN	1.4338	1.4667	1.3843	1.4581	**1.3684**	1.4426	1.4394	1.4301	1.4384	**1.4285**
ExpAlmon	1.3808	1.3808	1.3808	1.3808	1.3808	1.4446	1.4446	1.4446	1.4446	1.4446
Almon	1.4476	1.4705	1.4462	1.4189	1.4481	1.4647	1.4796	1.4418	1.4445	1.4439
Stepfun	1.4522	1.4632	1.4852	1.5207	1.5826	1.4525	1.4819	1.5087	1.5274	1.5383
UMIDAS	1.4508	1.5259	1.6819	1.6739	1.8077	1.4952	1.5500	1.6198	1.6644	1.6886
	2 lag orders for steam coal price					3 lag orders for steam coal L price				
Beta	1.4699	1.4544	1.4544	1.4544	1.4544	1.4386	1.4386	1.4386	1.4386	1.4386
BetaNN	1.4709	1.4542	**1.4459**	1.4540	1.4521	1.4628	1.4372	1.4306	1.4377	**1.4291**
ExpAlmon	1.4699	1.4699	1.4699	1.4699	1.4699	1.4492	1.4492	1.4492	1.4492	1.4492
Almon	1.4922	1.4972	1.4563	1.4568	1.4562	1.4802	1.4845	1.4463	1.4415	1.4429
Stepfun	1.4744	1.4990	1.5281	1.5414	1.5476	1.4608	1.4878	1.5160	1.5280	1.5371
UMIDAS	1.5175	1.5733	1.6375	1.6758	1.6932	1.5054	1.5574	1.6150	1.6643	1.6895
	4 lag orders for steam coal price					5 lag orders for steam coal price				
Beta	1.4090	1.4090	1.4090	1.4090	1.4090	1.4432	1.4432	1.4432	1.4432	1.4432
BetaNN	1.4513	1.4055	**1.3985**	1.4046	1.4063	1.4968	1.4417	**1.4340**	1.4420	1.4406
ExpAlmon	1.4189	1.4189	1.4189	1.4189	1.4189	1.4525	1.4525	1.4525	1.4525	1.4525
Almon	1.4470	1.4515	1.4146	1.4109	1.4098	1.4829	1.4981	1.4550	1.4557	1.4540
Stepfun	1.4268	1.4562	1.4899	1.5027	1.5130	1.4635	1.4943	1.5304	1.5465	1.5618
UMIDAS	1.4728	1.5273	1.5908	1.6382	1.6636	1.5109	1.5662	1.6297	1.6794	1.6992

Note: The bold values emphasize the smallest RMSEs of the single MIDAS model with different lag orders for steam coal price

Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy

the polynomial weight of BetaNN performs better in reflecting the relationship between the steam coal price and the crude oil price, and the best lag order for the crude oil price is 33. The empirical result shows that the best lag order for crude oil prices may be different from the variation of the lag order for steam coal price, which reflects the duration of the impact derived from the crude oil price to the steam coal price that ranges from 33 to 54 days. Meanwhile, BetaNN performs better in capturing the dynamic relationship between the steam coal price and the crude oil price than other polynomial weights in all conditions of different lag orders for the steam coal price and the crude oil price. What's more, when the best lag orders for the steam coal price and the crude oil price is 0 and 54, respectively, with the polynomial weight of BetaNN, namely, AR(0)-BetaNN-MIDAS(2, 54) show the highest predictive ability for the steam coal price.

Comparisons Among the Driving Factors

Considering the complex relationship between the steam coal price and the relevant daily factors, this paper selects the best polynomial weight, the best h step lead and the optimal lag orders for the steam coal price and the factors, respectively, to determine the best single MIDAS in the condition of different leads. Using the selecting mechanism introduced above, the best single MIDAS for the relevant factors are selected considering when various conditions, which is shown in Table 2. For crude oil, the best single MIDAS is still AR(0)-BetaNN-MIDAS(2, 54) with 1 step ahead, of which the polynomial weight is BetaNN, and the best lag order for the steam coal price and the crude oil price is 0 and 54, respectively. For natural gas, the best individual MIDAS with the highest predictive ability is chosen, of which the best lag order for the steam coal price and the natural gas price is 0 and 33 with the best weight Umidas when considering the condition of h = 1. For A-share, the best individual MIDAS with the highest predictive ability is chosen, of which the best lag order for the steam coal price and the A-share is 0 and 55 with the best weight Stepfun when considering the condition of h = 1. For temp, the best individual MIDAS with the highest predictive ability is chosen, of which the best lag order for the steam coal price and the temp is 4 and 36 with the best weight Stepfun when considering the condition of h = 4. For AQI, the best individual MIDAS with the highest predictive ability is chosen, of which the best lag order for the steam coal price and the AQI is 4 and 30 with the best weight Umidas when considering the condition of h = 4. The empirical result shows that the duration of the impact derived from crude oil price, A-share, temp, and natural gas to the steam coal price lasts longer than that of AQI. More specifically, the effect of A-share on the steam coal price lasts the longest among these daily factors while the AQI lasts the shortest. Natural gas price, AQI and temp are better predictors than A-share and crude oil price because the individual MIDAS with the former factors realizes the higher prediction accuracy. At the same time, considering the energy factors as the alternative to the steam coal, the effect of the crude oil lasts longer with the lower prediction accuracy while the natural gas shows a higher accuracy but lasting shorter Furthermore, when considering different leads of the steam coal price, natural gas price shows better predictive ability among these factors, followed by AQI, which indicating the reasonability of introducing it into the forecasting model.

Table 3 demonstrates the forecasting performance of the individual MIDAS-XGBoost with each factor after considering the error correction strategy. For crude oil, the best individual MIDAS-XGBoost is AR(0)-

Table 2. The best single MIDAS model with different leads for steam coal price.

Factors	The best MIDAS model	*RMSE*	The best MIDAS model	*RMSE*
	h = 1		h = 2	
Crude oil	**AR(0)-BetaNN-MIDAS(2, 54)**	**1.3684**	AR(0)-ExpAlmon-MIDAS(3, 38)	1.3691
Natural gas	**AR(0)-Umidas-MIDAS(6, 33)**	**1.2493**	AR(0)-Umidas-MIDAS(6, 32)	1.2693
A-share	**AR(0)-Stepfun-MIDAS(5, 55)**	**1.3362**	AR(0)-Almon-MIDAS(4, 46)	1.3505
Temp	AR(4)-Stepfun-MIDAS(5, 39)	1.3272	AR(4)-Stepfun-MIDAS(5, 29)	1.3380
AQI	AR(0)-Almon-MIDAS(4, 31)	1.3286	AR(4)-Umidas-MIDAS(6, 30)	1.3255
	h = 3		h = 4	
Crude oil	AR(0)-BetaNN-MIDAS(2, 59)	1.3802	AR(0)-BetaNN-MIDAS(2, 58)	1.3791
Natural gas	AR(0)-Umidas-MIDAS(6, 31)	1.2796	AR(0)-Umidas-MIDAS(6, 31)	1.3019
A-share	AR(0)-Almon-MIDAS(4, 45)	1.3559	AR(0)-Stepfun-MIDAS(5, 52)	1.3487
Temp	AR(4)-Umidas-MIDAS(6, 32)	1.3416	**AR(4)-Stepfun-MIDAS(5, 36)**	**1.3252**
AQI	AR(4)-Umidas-MIDAS(6, 31)	1.3173	**AR(4)-Umidas-MIDAS(6, 30)**	**1.3146**

Note: RMSE generated the single MIDAS model with each factor when considering different leads. The bold values emphasize the smallest RMSEs of the best individual model with each factor.

Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy

Table 3. The best single MIDAS model with different leads for steam coal price.

Factors	The best MIDAS model	*RMSE_COR*	The best MIDAS model	*RMSE-COR*
	h = 1		h = 2	
Crude oil	**AR(0)-BetaNN-MIDAS(2, 54)-XGB**	**0.3014**	AR(0)-ExpAlmon-MIDAS(3,38)-XGB	0.3101
Natural gas	AR(0)-Umidas-MIDAS(6, 33)-XGB	0.6247	**AR(0)-Umidas-MIDAS(6, 32)-XGB**	**0.6221**
A-share	AR(0)-Sepfun-MIDAS(5, 55)-XGB	0.4576	**AR(0)-Almon-MIDAS(4, 46)-XGB**	**0.2838**
Temp	AR(4)-Stepfun-MIDAS(5, 39)-XGB	0.9888	**AR(4)-Stepfun-MIDAS(5, 29)-XGB**	**0.9858**
AQI	**AR(0)-Almon-MIDAS(4, 31)-XGB**	**0.3050**	AR(4)-Umidas-MIDAS(6, 30)-XGB	0.9654
	h = 3		h = 4	
Crude oil	AR(0)-BetaNN-MIDAS(2, 59)-XGB	0.3083	AR(0)-BetaNN-MIDAS(2, 58)-XGB	0.3087
Natural gas	AR(0)-Umidas-MIDAS(6, 31)-XGB	0.6508	AR(0)-Umidas-MIDAS(6, 31)-XGB	0.6404
A-share	AR(0)-Almon-MIDAS(4, 45)-XGB	0.2871	AR(0)-Stepfun-MIDAS(5, 52)-XGB	0.4621
Temp	AR(4)-Umidas-MIDAS(6, 32)-XGB	1.0294	AR(4)-Stepfun-MIDAS(5, 36)-XGB	0.9893
AQI	AR(4)-Umidas-MIDAS(6, 31)-XGB	0.9610	AR(4)-Umidas-MIDAS(6, 30)-XGB	0.9561

Note: RMSE generated the single MIDAS model with each factor when considering different leads. The bold values emphasize the smallest RMSEs of the best individual model with each factor.

BetaNN-MIDAS(2, 54)-XGB for which the RMSE value is 0.3014 when considering the condition of h = 1. The forecasting accuracy of this hybrid model is approximately 77.46% greater than that of the single MIDAS with crude oil under similar conditions. For natural gas, the best individual MIDAS-XGBoost is AR(0)-Umidas-MIDAS(6, 32)-XGB for which the RMSE value is 0.6221 when considering the condition of h = 2. The forecasting accuracy of this hybrid model is approximately 50.20% greater than that of the single MIDAS with natural gas under similar conditions. For A-share, the best individual MIDAS-XGBoost is AR(0)-Almon-MIDAS(4, 46)-XGB for which the RMSE value is 0.2838 when considering the condition of h = 2. The forecasting accuracy of this hybrid model is approximately 78.76% greater than that of the single MIDAS with A-share under similar conditions. For temp, the best individual MIDAS-XGBoost is AR(4)-Stepfun-MIDAS(5, 29)-XGB for which the RMSE value is 0.9858 when considering the condition of h = 2. The forecasting accuracy of this hybrid model is approximately 25.61% greater than that of the single MIDAS with temp under similar conditions. For AQI, the best individual MIDAS-XGBoost is AR(0)-Almon-MIDAS(4, 31)-XGB for which the RMSE value is 0.3050 when considering the condition of h = 1. The forecasting accuracy of this hybrid model is approximately 76.79% greater than that of the single MIDAS with AQI under similar conditions. The empirical results show that the MIDAS-XGBoost model is generally better than the single MIDAS model with each factor, an improvement of approximately ranging 25.62% to 78.76%. The better performance of MIDAS-XGBoost indicates that there is a non-linearity in the steam coal price, and the addition of the XGBoost algorithm helps to capture this feature and improve the forecasting performance.

Evaluation of the Forecasts for Steam Coal Price

Forecast Comparison: Combination-MIDAS Model and Combination-MIDAS-XGBoost Model

The predictors set of the steam coal price in this paper show different predictive ability, for which the forecast performance is various under the different conditions discussed in section 4.2. To improve the stability of the predictive model under various conditions and achieve satisfactory performance, this paper constructs the combination-MIDAS model based on the best single model shown in Table 2. Instead of adopting one combination weight method to combine the forecast results of the best single model[48], five polynomial weight types are utilized in this paper as shown in Appendix A.2. The combination-MIDAS model can make full use of the information provided by the best single model with different factors. Thus it can address the misspecification bias and maintain the forecast ability under different conditions and structural breaks [41].

Table 4 demonstrates the out-of-sample forecast accuracy of the combination-MIDAS model. The empirical result shows that the predictive ability of the five weighted combination model is higher than the single MIDAS model, and this model maintains stability under different leads. In the condition of 0-step ahead, the combination-MIDAS model with DMSFE weight type reveals greater predictive power, followed by the MSFE and Equal weight types. In the condition of 1-step ahead, the combination -MIDAS model with DMSFE weight type reveals greater predictive power, followed by the MSFE and Equal weight types. Similarly, when considering the condition of 2-step, 3-step and 4-step ahead, the combination-MIDAS model with DMSFE

Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy

Table 4. RMSEs generated by the combination-MIDAS model.

Weights	h = 0	h = 1	h = 2	h = 3	h = 4
MSFE Type	1.2451	1.2318	1.2288	1.2251	1.2258
DMSFE Type	**1.2366**	**1.2248**	**1.2243**	**1.2204**	**1.2257**
AIC Type	1.3513	1.3272	1.3380	1.3416	1.3252
BIC Type	1.3513	1.3272	1.3380	1.3415	1.3252
Equal-Weights	1.2490	1.2356	1.2322	1.2282	1.2279

Note: The bold values emphasize the smallest RMSEs in the condition of different leads for the steam coal price.

Table 5. RMSEs generated by the combination-MIDAS-XGBoost model.

Weights	h = 0	h = 1	h = 2	h = 3	h = 4
MSFE Type	**0.2690**	**0.2832**	0.3019	**0.3083**	**0.3383**
DMSFE Type	0.2713	0.2998	**0.3004**	0.3091	0.3514
AIC Type	0.9908	0.9888	0.9858	1.0294	0.9893
BIC Type	0.9908	0.9888	0.9858	1.0293	0.9893
Equal-Weights	0.3509	0.3624	0.4901	0.4931	0.4989

Note: The bold values emphasize the smallest RMSEs in the condition of different leads for the steam coal price.

Table 6. Comparing of the best individual MIDAS model for each factor vs. AR, MA and ARMA model.

Weights	h = 0	h = 1	h = 2	h = 3	h = 4
Panel A: RMSE ratios of the best individual MIDAS model for each factor vs. AR(1)					
Crude oil	0.9182	0.9195	0.9200	0.9274	0.9267
Natural gas	0.8386	0.8395	0.8529	0.8598	0.8748
A-share	0.9089	0.8979	0.9075	0.9111	0.9063
Temp	0.9080	0.8918	0.8991	0.9015	0.8905
AQI	0.8903	0.8928	0.8907	0.8852	0.8833
Panel B: RMSE ratios of the best individual MIDAS model for each factor vs. MA(1)					
Crude oil	0.8647	0.8660	0.8664	0.8734	0.8727
Natural gas	0.7898	0.7906	0.8033	0.8098	0.8239
A-share	0.8560	0.8456	0.8546	0.8581	0.8535
Temp	0.8551	0.8399	0.8467	0.8490	0.8386
AQI	0.8385	0.8408	0.8388	0.8336	0.8319
Panel C: RMSE ratios of the best individual MIDAS model for each factor vs. ARMA(1,1)					
Crude oil	0.9218	0.9232	0.9236	0.9311	0.9304
Natural gas	0.8419	0.8428	0.8563	0.8633	0.8783
A-share	0.9125	0.9014	0.9111	0.9147	0.9099
Temp	0.9116	0.8954	0.9027	0.9051	0.8940
AQI	0.8939	0.8963	0.8942	0.8887	0.8869

Note: If the RMSE ratios are less than 1, that indicating the prediction accuracy of the best individual MIDAS model for each factor are better than the benchmark models, e.g., AR(1), MA(1), ARMA(1,1).

Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy

weight type reveals greater predictive power, followed by the MSFE and Equal weight types.

In addition, the predictive performance of AIC weight type and BIC weight type are not satisfying when considering each step. Thus, the DMAFE has the ability to extracting the predictive information from the optimal individual MIDAS model, and the forecast robustness of the DMSFE weight type outperforms the other alternative weight types.

Table 5 demonstrates the predictive RMSE value of the combination-MIDAS-XGBoost model when considering different leads of the steam coal price. It's clear from the empirical results that the out-of-sample forecast accuracy of the proposed hybrid model is higher than that of the traditional combination-MIDAS. When considering the condition of 0-step ahead, the MSFE weight type shows the highest predictive power among these weight types and the RMSE value is 0.2690. Comparing the forecast accuracy of the combination-MIDAS with the MSFE weight type, modified combination MIDAS with 0 lead by the XGBoost is approximately 78.39% higher in predictive power. Similarly, when considering h-step leads is 1, 2, 3, and 4, respectively, the proposed hybrid model with MSFE weight type shows the highest predictive power and forecast stability. Therefore, the XGBoost algorithm can improve the forecast performance of the individual MIIDAS and the combination-MIDAS model by capturing the nonlinear trends of steam coal price.

Forecast Comparison: Individual-MIDAS Models vs. AR, MA, and ARMA

What's more, the paper compares the forecast performance of the best individual MIDAS model with that of the benchmark models, e.g., AR, MA, and ARMA, to emphasize the predictive power of the MIDAS regression model. Table 6 shows the RMSE ratios of the best individual MIDAS model for each factor vs. that of AR(1), MA(1), and ARMAR(1,1). The empirical result shows that the forecast performance of the best individual MIDAS model for each factor is higher approximately from 7.89% to 21.02% compared with these traditional time series prediction models listed above. Taking the RMSE ratios shown in Panel A of Table 6 as an example, when considering the condition of 0-step ahead, the best individual MIDAS-XGBoost model with natural gas shows the greater predictive power and the forecast accuracy of this model is higher approximately 16.14% than that of AR(1). Similarly, when considering the condition of 1-step or 2-step ahead, the forecast accuracy of the best individual MIDAS model with natural gas is higher approximately 16.05% and 15.71% than that of AR(1), respectively. Generally, the daily predictor could provide rich forecast information for the steam coal price prediction, thus enhance the forecast accuracy of the predictive model.

Conclusions

This paper proposes combination-MIDAS-XGBoost models for the weekly steam price forecast in China with the daily various factors obtained. To realize the accurate forecast and improve the predictive stability, the proposed hybrid model takes advantage of the individual MIDAS model and the XGB algorithm. First, the individual MIDAS model utilizes the daily frequency factors for coal price forecast, including energy factors, economic factors, weather, and environment. More specifically, the crude oil price and the natural gas from the energy market, the A-share index from the economic market, the temp from the weather aspect and the AQI from the environmental aspect. Then, the best individual MIDAS model is selected by comparing the RMSE values of this model with different factors. Second, the XGBoost algorithm can capture the nonlinear characters of steam coal price and thus eliminate the uncertainty faced by the individual MIDAS model. Third, five weight types are utilized to construct the combination forecast model, which makes use full of forecast information of the individual models, thus address the misspecification bias and improve the predictive power and stability.

Some conclusions are drawn in this paper according to the empirical results.

The impact of crude oil price, A-share, temp, and natural gas to the steam coal price is more sustained than that of AQI. At the same time, energy and environmental factors are more predictive for the steam coal price than economics and weather. The steam price shows significantly auto-correlative in the energy market of China, lasting from 0 to 4 weeks. XGBoost algorithm is effective in modifying the individual MIDAS model under different conditions. At the same time, the forecast accuracy of the MIDAS-XGBoost hybrid model increases approximately ranging from 28.61% to 78.76%, which is better than the single MIDAS with different factors. When constructing the combination-MIDAS model, the forecast robustness of the DMSFE weight type outperforms the other alternative weight types. The out-of-sample forecast accuracy of the combination-MIDAS-XGBoost hybrid model is higher than that of the benchmark models and the combination-MIDAS, an improvement of approximately 23.27% and 78.39%, respectively.

There are still some potential extensions to our research in the future. For example, although we focused on the weekly steam coal price forecast, it would have been straightforward to extend our analysis to the daily horizon with higher frequency predictors, such as the hourly inter-data to perform real-time forecast, if they are available. One could also extend our analysis to choose another predictive model to correct the error of the combination-MIDAS. Doing so would raise additional applications, and give more stable forecast performance in an uncertainty context.

Acknowledgements

This work is supported by the National Science Foundation of China (No. 71973132); Taishan Scholar Program (tsqn20161014; ts201712014).

Conflicts of Interest

The authors declare no conflict of interest.

Appendix A

A.1 Combination weights

(i) MSFE-weighted type

MSFE refers to the squared forecast error, which is used to combine the individual approach. The weight is given as:

$$w_{j,\mathrm{T}} = m_{j,T}^{-1} / \sum_{j=1}^{n} m_{j,T}^{-1} \tag{A.1}$$

...where $m_{j,T} = \sum_{i=T_0}^{t} (\delta^{i-T_0} (y_{j,T+s} - \hat{y}_{j,T+s|T}))^2 / (t - T_0 + 1)$.

When δ = 1, $m_{j,T}$ is defined as the MSFE of the individual MIDAS-XGBoost with jth factor. Here, $t - T_0 + 1$ indicates the number of observations in the out-of-sample, and $y_{j,T+s}$ refers to the real observation.

(ii) DMSFE-weighted type

When δ = 0.9 in the equation of MSFE discussed above, it refers to the discounted squared forecast error (DMSE)-weighted type.

(iii) AIC-weighted type

AIC refers to Akaike information criteria, and AIC-weighted type is defined as:

$$w_{j,T} = exp(-AIC_j) / \sum_{j=1}^{N} exp(-AIC_j) \tag{A.2}$$

(vi) BIC-weighted type

BIC refers to Bayesian information criteria, and BIC-weighted type is defined as:

$$w_{j,T} = exp(-BIC_j) / \sum_{j=1}^{N} exp(-BIC_j) \tag{A.3}$$

(v) Equal-weighted type

The equal-weighted type refers to the average weight, as is simply given:

$$w_{j,T} = 1 / N \tag{A.4}$$

A.2 Polynomial weight functions

(i) The beta density function polynomial is selected for constructing two forms of weight functions including the beta density function polynomial with zero lag (Beta) and the beta density function polynomial with non-zero lag (BetaNN), which are defined as:

$$w(k;\theta) = w(k;\theta_1,\theta_2,\theta_3) = \frac{f(x_i,\theta_1,\theta_2)}{\sum_{k=1}^{K} f(x_i,\theta_1,\theta_2)} + \theta_3 \tag{A.5}$$

...where $x_i = k/K$, $f(x_i,\theta_1,\theta_2) = x_i^{\theta_1-1}(1-x_i)^{\theta_2-1}\Gamma(\theta_1+\theta_2)/\Gamma(\theta_1)\Gamma(\theta_2)$, and $\Gamma(\theta) = \int_0^{\infty} e^{-x}x^{\theta-1}dx$.

When $\theta_3 = 0$, $w(k;\theta) = w(k;\theta_1,\theta_2) = f(x_i,\theta_1,\theta_2) / \sum_{k=1}^{K} f(x_i,\theta_1,\theta_2)$. Here, $w(k;\theta)$ refers to the beta polynomial (Beta). When $\theta_1 = 0$, $w(k;\theta) = w(k;1,\theta_2,\theta_3) = f(x_i,1,\theta_2) / \sum_{k=1}^{K} f(x_i,1,\theta_2) + \theta_3$. Here, $w(k;\theta)$ refers to the beta polynomial with non-zero lag (BetaNN).

(ii) The exponential Almon lag polynomial of order p (ExpAlmon) is defined as:

$$w(k;\theta) = w(k;\theta_1,\theta_2,\cdots,\theta_p) = \frac{e^{\left(\theta_1 k^1 + \theta_2 k^2 + \cdots + \theta_p k^p\right)}}{\sum_{k=1}^{K} e^{\left(\theta_1 k^1 + \theta_2 k^2 + \cdots + \theta_p k^p\right)}} \tag{A.6}$$

(iii) The Almon lag polynomial of order p (Almon) is defined as:

$$w(k;\theta) = w(k;\theta_1,\theta_2,\cdots,\theta_p) = \sum_{p=0}^{P} \theta_p k^p \tag{A.7}$$

(vi) Polynomial specification with step function (Stepfun) is defined as:

$$w(k;\theta) = w(k;\theta_1,\theta_2,\cdots,\theta_p) = \theta_1 I_{i\in[\alpha_0,\alpha_1]} + \sum_{p=2}^{P} \theta_p I_{i\in[\alpha_{p-1},\alpha_p]}$$

$$I_{i\in[\alpha_{p-1},\alpha_p]} = \begin{cases} 1, & \alpha_{p-1} < i < \alpha_p \\ 0, & otherwise \end{cases} \tag{A.8}$$

...where $\alpha_0 = 1 < \alpha_1 < \alpha_2 < ... < \alpha_p < K$.

(v) Unrestricted MIDAS (UMIDAS) with the autoregressive items generalizes to:

$$\phi(L)Y_t = \alpha + \beta(L)\mathrm{X}_t^{(m)} + \mu_t \tag{A.9}$$

...where $\beta(L) = \sum_{k=0}^{K} \beta_k L^{k/m}$, and $\phi(L) = 1 - \sum_{p=1}^{P} \varphi_k L^p$.

Author Contributions

Conceptualization, Chunyang Wang and Wanglin Kang; Data curation, Chunyang Wang; Formal analysis, Chunyang Wang; Investigation, Wanglin Kang; Methodology, Chunyang Wang; Project administration, Chunyang Wang; Resources, Wanglin Kang; Software, Wanglin Kang; Supervision, Chunyang Wang; Validation, Wanglin Kang; Writing – original draft, Chunyang Wang; Writing – review and editing, Wanglin Kang.

Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy

Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy

References

1. LI J., XIE C., LONG H. The roles of inter-fuel substitution and inter-market contagion in driving energy prices: Evidences from China's coal market. Energy Economics, **84**, 104525, **2019**.
2. WEINBERGER G., AMIRI S., MOSHFEGH B. On the benefit of integration of a district heating system with industrial excess heat: An economic and environmental analysis. Applied Energy, **191**, 454, **2017**.
3. ZHANG Q., LI Y., XU J., JIA G. Carbon element flow analysis and CO_2 emission reduction in iron and steel works. Journal of Cleaner Production, **172**, 709, **2018**.
4. PAPIEŻ M., ŚMIECH S. Dynamic steam coal market integration: Evidence from rolling cointegration analysis. Energy Economics, **51**, 510, **2015**.
5. DING Z., FENG C., LIU Z., WANG G., HE L., LIU M. Coal price fluctuation mechanism in china based on system dynamics model. Natural Hazards, **85** (2), 1151, **2017**.
6. KRZEMIEŃ A., FERNÁNDEZ P.R., SÁNCHEZ A.S., LASHERAS F.S. Forecasting European thermal coal spot prices. Journal of Sustainable Mining, **14** (4), 203, **2015**.
7. MOKNI K. A dynamic quantile regression model for the relationship between oil price and stock markets in oil-importing and oil-exporting countries. Energy, 118639, **2020**.
8. PAPIEŻ M., ŚMIECH S. The analysis of relations between primary fuel prices on the European market in the period 2001-2011. Rynek Energ II, **5** (96), 139, **2011**.
9. LINN J., MCCORMACK K. The roles of energy markets and environmental regulation in reducing coal-fired plant profits and electricity sector emissions. The RAND Journal of Economics, **50** (4), 733, **2019**.
10. DOYLE M., FELL H. Fuel prices, restructuring, and natural gas plant operations. Resource and Energy Economics, **52**, 153, **2018**.
11. LI J., XIE C., LONG H. The roles of inter-fuel substitution and inter-market contagion in driving energy prices: Evidences from China's coal market. Energy Economics, **84**, 104525,**2019**.
12. GU F., WANG J., GUO J., FAN Y. How the supply and demand of steam coal affect the investment in clean energy industry? Evidence from China. Resources Policy, **69**, 101788, **2020**.
13. LIU X., GUO P., GUO S, Assessing the eco-efficiency of a circular economy system in China's coal mining areas: Emergy and data envelopment analysis. Journal of Cleaner Production, **206**, 1101, **2018**.
14. SADHU M., GOSWAMI U., DAS N., SADHU P.K., GOSWAMI A. Improvement of energy forecasting model to safeguard energy security in India. Journal of Renewable and Sustainable Energy, **10** (6), 065907, **2018**.
15. SRIKANTH R. India's sustainable development goals – Glide path for India's power sector. Energy policy, **123**, 325, **2018**.
16. SHAHBAZ M., KHAN S., TAHIR M.I. The dynamic links between energy consumption, economic growth, financial development and trade in China: fresh evidence from multivariate framework analysis. Energy economics, **40**, 8, **2013**.
17. ZHANG B., MA J. Coal price index forecast by a new partial least-squares regression. Procedia Engineering, **15**, 5025, **2011**.
18. KRZEMIEŃ A., FERNÁNDEZ P.R., SÁNCHEZ A.S., LASHERAS F.S. Forecasting European thermal coal spot prices. Journal of Sustainable Mining, **14** (4), 203, **2015**.
19. ZHAO Z.Y., ZHU J., XIA B. Multi-fractal fluctuation features of thermal power coal price in China. Energy, **117**, 10, **2016**.
20. FAN X., WANG L., LI S. Predicting chaotic coal prices using a multi-layer perceptron network model. Resources Policy, **50**, 86, **2016**.
21. ALAMEER Z., FATHALLA A., LI K., YE H., JIANHUA Z. Multistep-ahead forecasting of coal prices using a hybrid deep learning model. Resources Policy, **65**, 101588, **2020**.
22. WANG Q., LI S., LI R. Will Trump's coal revival plan work?-Comparison of results based on the optimal combined forecasting technique and an extended IPAT forecasting technique. Energy, **169**, 762, **2019**.
23. WANG C., LI B.B., LIANG Q.M., WANG J.C. Has China's coal consumption already peaked? A demand-side analysis based on hybrid prediction models. Energy, **162**, 272, **2018**.
24. GHYSELS E., SANTA-CLARA P., VALKANOV R. The MIDAS touch: Mixed data sampling regression models, **5** (1), 512, **2004**.
25. BAUMEISTER C., GUÉRIN P., KILIAN L. Do high-frequency financial data help forecast oil prices? The MIDAS touch at work. International Journal of Forecasting, **31** (2), 238, **2015**.
26. ANDREOU E., GHYSELS E., KOURTELLOS A. Should macroeconomic forecasters use daily financial data and how? Journal of Business & Economic Statistics, **31** (2), 240, **2013**.
27. BAUMEISTER C., KILIAN L. Lower oil prices and the US economy: Is this time different? Brookings Papers on Economic Activity, **2**, 287, **2016**.
28. ANAND A., SUGANTHI L. Forecasting of electricity demand by hybrid ANN-PSO models. International Journal of Energy Optimization & Engineering, **6** (4), 66e83, **2017**.
29. CHEN T., GUESTRIN C. Xgboost: A scalable tree boosting system. In Proceedings of the 2^{2nd} ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, **13** (17), 785, **2016**.
30. LI G.Y., LI W., TIAN X.L., CHE Y.F. Short-term electricity load forecasting based on the XGBoost algorithm. Smart Grid Smart Grid, **7** (4), 274, **2017**.
31. SALISU A., GUPTA R. Oil shocks and stock market volatility of the BRICS: A GARCH-MIDAS approach. Global Finance Journal, **1**, 100546, **2020**.
32. FANG T., LEE T., SU Z. Predicting the long-term stock market volatility: A GARCH-MIDAS model with variable selection. Journal of Empirical Finance, **58**, 36, **2020**.
33. LU X., MA F., WANG J., WANG J., Examining the predictive information of CBOE OVX on China's oil futures volatility: Evidence from MS-MIDAS models. Energy, **212** (1), 118743, **2020**.
34. BAUMEISTER C., GUÉRIN P., KILIAN L. Do high-frequency financial data help forecast oil prices? The MIDAS touch at work. International Journal of Forecasting, **31** (2), 238, **2015**.
35. JIANG Y., GUO Y., ZHANG Y. Forecasting China's GDP growth using dynamic factors and mixed-frequency data. Economic Modelling, **66**, 132, **2017**.
36. QIU Y. Forecasting the Consumer Confidence Index with tree-based MIDAS regressions. Economic Modelling, **91**, 247, **2020**.

37. ZHAO X., HAN M., DING L., KANG W. Usefulness of economic and energy data at different frequencies for carbon price forecasting in the EU ETS. Applied Energy, **216**, 132, **2018**.
38. ZHAO, X., HAN, M., DING, L., CALIN, A. C. Forecasting carbon dioxide emissions based on a hybrid of mixed data sampling regression model and back propagation neural network in the USA. Environmental Science and Pollution Research, **25** (3), 2899, **2018**.
39. YOU Y., LIU X. Forecasting short-run exchange rate volatility with monetary fundamentals: A GARCH-MIDAS approach. Journal of Banking & Finance, **116**, 105849, **2020**.
40. ZHANG Y.J., WANG J.L. Do high-frequency stock market data help forecast crude oil prices? Evidence from the MIDAS models. Energy Economics, **78**, 192, **2009**.
41. HAN M., DING L., ZHAO X., KANG W. Forecasting carbon prices in the Shenzhen market, China: The role of mixed-frequency factors. Energy, **171**, 69, **2019**.
42. DING L., LV Z., HAN M., ZHAO X., WANG W. Forecasting China‘s wastewater discharge using dynamic factors and mixed-frequency data. Environmental Pollution, **255**, 113148, **2019**.
43. LI P., ZHANG J. A new hybrid method for China’s energy supply security forecasting based on arima and xgboost. Energies, **1** (7), 1687, **2018**.
44. HUANG Y., YEN M. A new perspective of performance comparison among machine learning algorithms for financial distress prediction. Applied Soft Computing, **83**, 105663, **2019**.
45. WANG C., GUO J . A data-driven framework for learners‘ cognitive load detection using ECG-PPG physiological feature fusion and XGBoost classification. Procedia Computer Science, **147**, 338, **2019**.
46. MA X., SHA J., WANG D., YU Y., YANG Q., NIU X. Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning. Electronic Commerce Research and Applications, **31**, 24, **2018**.
47. BAI J., GHYSELS E., WRIGHT J.H. State space models and MIDAS regressions. Econometric Reviews, **32** (7), 779, **2013**.
48. GHYSELS E., OZKAN N. Real-time forecasting of the US federal government budget: A simple mixed frequency data regression approach. International Journal of Forecasting, **31** (4), 1009, **2015**.

Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy • Author Copy