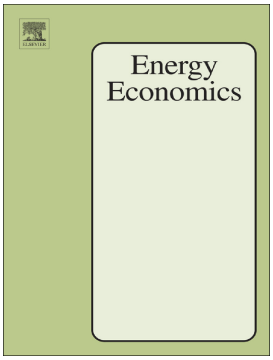

Coal overcapacity in China: Multiscale analysis and prediction

Delu Wang, Yadong Wang, Xuefeng Song, Yun Liu

PII: S0140-9883(18)30012-4
DOI: https://doi.org/10.1016/j.eneco.2018.01.004
Reference: ENEECO 3878

To appear in:

Received date: 5 June 2017
Revised date: 15 November 2017
Accepted date: 3 January 2018

Please cite this article as: Delu Wang, Yadong Wang, Xuefeng Song, Yun Liu , Coal overcapacity in China: Multiscale analysis and prediction. The address for the corresponding author was captured as affiliation for all authors. Please check if appropriate. Eneeco(2018), https://doi.org/10.1016/j.eneco.2018.01.004

Coal overcapacity in China: Multiscale analysis and prediction

Delu Wang[a, 1], Yadong Wang[a], Xuefeng Song[b], Yun Liu[c]

[a] School of Management, China University of Mining and Technology, Xuzhou, Jiangsu 221116, P. R. China

[b] School of Management Science and Industrial Engineering, Nanjing University of Finance and Economics, Nanjing, Jiangsu 210003, P. R. China

[c] Management School, Lancaster University, Lancaster LA2 0PJ, United Kingdom

Abstract

Gaining a quantitative understanding of the causes of coal overcapacity and accurately predicting it are important for both government agencies and coal enterprises. Following the decomposition-reconstruction-prediction concept, a combined Ensemble Empirical Mode Decomposition-Least Square Support Vector Machine-Autoregressive Integrated Moving Average (EEMD-LSSVM-ARIMA) model is proposed for quantitatively analyzing and forecasting coal overcapacity in China. The results show that the main causes of coal overcapacity in China include insufficient demand, market failure, and institutional distortion. Institutional distortion, with an influence degree of 73.75%, is the most fundamental and influential factor. From 2017 to 2019, the scale of coal overcapacity in China will reach between 1.721and 1.819 billion tons, suggesting that coal overcapacity will remain a serious problem. The rate of coal overcapacity caused by insufficient demand will fluctuate slightly, while coal overcapacity caused by market failure will trend downward, but the impact of institutional distortion on coal overcapacity will be exacerbated. A statistical analysis demonstrates that the EEMD-LSSVM-ARIMA model significantly outperformed other widely developed baselines (e.g. ARIMA, LSSVM, EEMD-ARIMA, and EEMD-LSSVM).

Keywords: coal industry; overcapacity; multiscale analysis; forecasting; integrated model

Jel classification: C53; Q41; Q48; L72

[1] Corresponding author.

E-mail address: dlwang@cumt.edu.cn (D.L. Wang)

1. Introduction

Scientific planning for the production capacity of coal, a basic energy resource, is not only critical for the healthy development of the coal industry but is also important for national energy security and promoting the sustainable development of the national economy (Tang and Peng, 2017). Li et al. (2015) stated that approximately 77% of energy in China was generated by the coal industry. Many enterprises have been competing in the coal industry since the beginning of the 21st century, when the sustained and rapid growth of China's macro economy began and triggered continuous increases in coal prices. Coal output has continued to grow, and China's coal industry has enjoyed a 10-year "golden period" of development. However, the coal overcapacity problem is becoming more serious due to the global economic downturn, market failure, institutional distortions, and energy transformation, and coal prices have been decreasing since 2012 (Tang et al., 2016; Tang and Peng, 2017; Wang et al., 2017). A certain amount of excess coal capacity is normal for a market economy; if there were no excess capacity, the market could not compete, and it would be difficult to achieve "survival of the fittest" and effective economic development. Once coal overcapacity becomes a serious problem, however, grave consequences can ensue if effective measures are not taken, such as economic fluctuations, vicious market competition, serious wastage of resources, decline in corporate profits, product price distortions, and environmental pollution, impeding the healthy and harmonious development of the national economy (Yuan et al., 2016; Zhang et al., 2016). The overcapacity problem has been a longstanding "chronic illness" affecting China's industrial development, and a successful resolution of surplus coal capacity is key for the sustainable development of China's economy.

The Chinese government has recently developed a series of policies to resolve overcapacity in the coal industry and other industries. In 2009, the State Council issued the *Circular of the State Council on Several Opinions on Restraining Overcapacity and Repeated Construction of Certain Industries and Guiding the Healthy Development of the Industry by the State Development and Reform Commission.* In 2013, the State

Council issued the *Guidance on Resolving the Contradictions of Serious Overcapacity.* In February 2016, the State Council's *Opinions of the State Council on Resolving the Overcapacity Problem of the Coal Industry to Realize Development by Extricating the Coal Industry from Difficulties* stipulated that, by 2020, the coal industry must have reduced its production capacity by 500 million tons and restructured a further 500 million tons of outdated production capacity. The National Energy Work Conference held in December of the same year emphasized that, over the course of China's 13th Five-year Plan, the proportion of coal consumption must be further reduced and proposed withdrawing 800 million tons of coal production capacity. The Central Economic Work Conferences, which have been held since 2012, have stated that the core task of China's supply-side reform is de-capacity, whereby valuable resource elements are released from industries whose growth is limited to ease the pressure of surplus capacity. The implementation of related policy measures has not yet been successful. Excess capacity has not been effectively suppressed but has entered a cycle of "surplus–regulation–over-surplus–over-regulation." The overcapacity problem has found no solution, even after extensive regulation, leading to a "more control-more chaos" situation. Excess capacity shows a worsening trend and has evolved into a critical, wide-ranging, and chronic overcapacity after being a manageable, contained, and underlying problem. From a local perspective, the de-capacity schedules of each region in China are inconsistent: some cities and provinces are progressing slowly, and many regions are achieving their de-capacity targets through the fraudulent practice of "approved production capacity reduction." From a narrower perspective, many enterprises, including those that are large and state-owned, are self-deceptive when implementing de-capacity measures. A field survey[2] on the de-capacity results of 12 mining groups in Yulin, Ordos, and other key coal-production areas found that, while some enterprises were closing small and medium-sized mines with

[2] To better understand the current situation, characteristics, and causes of overcapacity in China's coal industry as well as the main obstacles to and difficulties of de-capacity, the authors carried out field research in coal fields such as those in Erdos, Datong, Yulin, Jining, and other key coal cities from March to July 2010 and conducted in-depth interviews with relevant government departments and coal enterprises.

production capacities of less than 600,000 tons, they also rebuilt and expanded larger mines through technological transformation. The increased capacity from these large mines was much higher than the capacity conserved by closing smaller mines.

Given the serious impacts of overcapacity on sustainable economic development, many scholars have attempted to understand the causes and mechanisms that lead to overcapacity (Wang et al., 2014; Dagdeviren, 2016) and develop measurement methods (Klein and Preston, 1967; Ray, 2015; Arfa et al., 2017) and governance policies (Wu and Li, 2015; Zhang, 2016). Quantitative studies on the causes and prediction of coal overcapacity are limited. The first step in effectively controlling the coal overcapacity problem is to accurately identify its causes and future trends and then build an overcapacity policy system. We conduct an exploratory study on the causes and prediction of coal overcapacity. This study contributes to the literature in three ways. First, an EEMD-LSSVM-ARIMA method is proposed that, by integrating three models, benefits from their individual advantages and avoids their disadvantages. This new hybrid model is proven to be a practical and effective tool for analyzing and forecasting coal overcapacity in China. Second, most of the existing research on the causes of excess capacity is qualitative and thus cannot quantify the impacts of the relevant overcapacity drivers. Through EEMD processing, this study performs a multiscale analysis on the three main causes of Chinese coal overcapacity and their degrees of influence, finding that institutional distortion is the fundamental determinant. This finding should help policymakers create targeted government programs. Third, most of the predictions for the coal industry concern coal demand and output; predictions about coal overcapacity are scant. We forecast the evolution of coal overcapacity and find that, from 2017 to 2019, the overall size of the overcapacity will fluctuate between 1.721 billion tons and 1.819 billion tons, and that the effect of institutional distortions on overcapacity will become more severe, offering important implications for coal industry participants and investors.

The remainder of this paper is structured as follows. Section 2 reviews the literature. Section 3 presents the

integrated EEMD-LSSVM-ARIMA methodology for the analysis and prediction of coal overcapacity. Section 4 describes the study's sample data. Section 5 reports the empirical results and discusses them. Finally, section 6 summarizes the study's key conclusions and policy implications.

2. Literature review

2.1. Formation mechanism of coal overcapacity

On an industrial level, overcapacity occurs when actual industrial output is lower than the industrial production capacity during a certain period of time. Scholars have widely studied the mechanisms of the formation of overcapacity as an economic phenomenon from several perspectives. In Western countries, overcapacity is most often a short-term economic event accompanying an economic crisis (Greenblatt, 2017). Research on this issue has been conducted from three main perspectives in attempts to explain the mechanism of overcapacity formation on the basis of micro theory. The first perspective is that of game theory, focusing on how the decisions made by leading enterprises facing the threat of invasion from potential competitors affect overcapacity (Mathis and Koscianski, 1997). The second is the perspective of oligarchy and market characteristics in terms of game theory, which focuses on how investment and price strategies undertaken by enterprises aiming for benefit maximization result in excess capacity (Davidson and Deneckere, 1990). The last perspective is that of future market uncertainty, which focuses on exploring how investment decisions made by enterprises aiming to enhance the value of "operational options" lead to surplus capacity (Pindyck, 1986). Some scholars have also analyzed the microscopic mechanisms of overcapacity from such perspectives as structural constraints (Dagdeviren, 2016), business cycles (Mulligan, 2016), and price mechanism (Pirard, 2007).

Other scholars have researched the mechanisms of Chinese overcapacity in terms of two core perspectives, market failure and institutional distortion, to understand why Chinese overcapacity differs from that in Western countries and why both cyclical and non-cyclical overcapacity occurs. The market failure hypothesis posits that overcapacity results mainly from the market economy itself and is represented by the investment "wave

phenomenon" proposed by Lin (2010). In developing countries, the wave phenomenon is caused by an "investment surge" caused by a social consensus on the prospects for an industry. Due to the acceleration of industrialization and urbanization in China, Chinese society has a strong positive perception of the coal, steel, and other industries. A large amount of social investment is flowing into several major industries, and competing investments are being made in constructing projects. The capacity scale has increased rapidly, leading to rapid development in the coal industry and unprecedented industrial expansion (Lin et al., 2010). The institutional distortion hypothesis argues that the root of Chinese overcapacity is an imbalance between the relationships involved in the process of economy transition. China's fiscal decentralization is exacerbating competition between local governments; to maximize GDP and tax revenues, they use various methods to attract investment, which lead to severe market segmentation, excessive investment, and homogeneous industrial structures, and thus to overcapacity (Jiang et al., 2012). The superposition effects of ineffective central and local government intervention are also important causes of excess capacity in China (Zhang et al., 2017).

The research perspectives and opinions on overcapacity formation mechanisms vary. Studies of the causes of overcapacity in Western countries are based on market determination of resource allocation. The small-scale mechanisms of overcapacity are primarily related to market competition between enterprises; however, because of the strong and comprehensive influence exerted by Chinese industrial development policies as well as excessive state intervention and administrative management, the fundamental causes of overcapacity in China differ greatly from those in Western countries. This may be related to the inconsistency of resource allocation under the Chinese system. Most studies perform qualitative analyses of the causes of overcapacity in China and fail to examine the extent to which the causes have influenced the volatility of industrial overcapacity, making it difficult to provide adequate guidance to policymakers.

2.2. Measurement and forecast of coal overcapacity

The measurement and forecasting of overcapacity scales, two core concerns in the quantitative research on coal overcapacity, are prerequisites for effective targeted governance policies. Most studies have used capacity utilization as an indicator of overcapacity. Other measurement methods include the peak value method (Klein and Preston, 1967), the production function method (Zhang et al., 2016), the cost function method, the cointegration analysis method (Ray, 2015), data envelopment analysis (Karagiannis, 2015), and stochastic frontier analysis (Arfa et al., 2017). The basic procedure of these methods is to first estimate the output of production capacity and then use actual output divided by production capacity to calculate capacity utilization. The cost function method can account for the various input factors consumed during the production process and can calculate production costs through the prices of the input factors, which is not possible with other calculation methods. This method has been widely used. Utilization reflects the overall level of overcapacity in a particular industry, but it cannot quantitatively reflect the driving factors of overcapacity (i.e., the contribution made by individual factors to overcapacity volatility).

Reliable forecasting of industrial overcapacity, an important reference for decision-making, can provide effective assistance in the development of national overcapacity governance policies. Forecasts of overcapacity are scarce relative to the abundant research on overcapacity measurement. Several energy prediction research scholars have explored energy production, consumption, and demand forecasts, including those for coal, oil, and shale gas (Yu et al, 2012; Wang et al., 2016; Liu et al., 2017). The methods and models used in these studies can be divided into five categories: regression models, the gray prediction method, the system dynamics method, artificial intelligence algorithms, and combination models (Suganthi and Samuel, 2012). Each model has specific applications according to its mathematical theory. These prediction models have unique advantages but also corresponding drawbacks. The application scope, advantages, and disadvantages of each model are summarized in Table 1.

Table 1

Typical literature and models for forecasting energy production and consumption

Forecasting model	Typical literature	Advantages	Disadvantages
ARMA model	Ediger and Akar, 2007; Wang et al., 2012;	When modeling non-stationary series, stochastic perturbation factors are well-handled.	Important information in the original data are often lost in the modification process.
Regression model	Höök and Aleklett, 2009; Höök et al., 2010; Aydin, 2015	The model parameter estimation technique is relatively well-developed, and the prediction process is simple.	It requires that the data meet certain prerequisites, and the regression variables are not well-reflected; it assumes uniform impacts of the data on the forecast, which is not realistic.
Gray prediction model	Hu, 2017; Xie et al., 2015; Tsai, 2016	Data for the influential factors on the forecast object are not required, which can overcome the issues caused by a lack of historical data and low reliability.	The higher the gray scale of the data, the lower the prediction accuracy; it is unsuitable for series that have an exponential trend.
System dynamics model	Feng et al., 2013; Wang et al., 2017	It can reflect the impacts of multiple factors on the predicted object and identify nonlinear, dynamic relationships between them.	It is difficult to accurately establish nonlinear and dynamic relationships between the influential factors and runs at a high cost.
Neural network	Hu and Zhao, 2008; Wu et al., 2013	It has the adaptive functions of a series of non-structural, non-precision laws; it does not need to calculate statistical characteristics when modeling.	The algorithm easily falls into minimum local points, the determination of the structure type is too dependent on experience, its training speed is slow, and it also easily over-fits the data.
Support vector machine	Jain et al., 2014; Papadimitriou et al., 2014	It can solve common problems of pattern recognition such as small samples, nonlinearity, high dimension, and local minimum.	It is sensitive to missing data; the training of large-scale samples is difficult.

Single models are easy to operate, but they lack comprehensive awareness of the basic laws of energy supply and demand; moreover, the model's character determines its imperfect prediction information, so its forecast accuracy needs to be improved. For example, a basic assumption of the autocorrelation time series model is that all of the factors affecting changes in energy production have been reflected in the variable itself, so future yield can be inferred from historical production. This method can eliminate random fluctuations in the sequence and fit a qualitative energy demand trend, but it cannot reflect the intricate relationships between the complex energy demand systems. The combined model aggregates information from various data samples and the useful aspects of different models and can thus account for more comprehensive information than a single prediction model can. The combined model can thus effectively reduce or overcome the negative impacts of some of the factors in single models and thus improve prediction accuracy and credibility. Existing combination

models include the ARMA-GARCH (Liu and Shi, 2013), GM-ARMA (Xu et al., 2015), EEMD- LSSVM-GARCH (Zhang, et al., 2015), and Wavelet transform-ARMA-Learning machine (Yang et al., 2017).

3. Methodology

3.1. Framework

Based on decomposition-reconstruction-prediction methods, the EEMD-LSSVM-ARIMA model was produced to analyze and forecast overcapacity volatility in China's coal industry. There are four main steps in the proposed methodology. First, to identify the deep influential factors, we extended the EEMD using a mirror extending method to decompose the coal overcapacity data series and obtained intrinsic mode functions (IMFs) and one residual (RES) with different frequencies. Second, to emphasize the economic value of the decomposition results and reduce the computational cost of prediction, we reconstructed decomposed modes with similar data characteristics into a high-frequency component (HFC), low frequency component (LFC), and a trend component (RES). Third, the complementary ARIMA and LSSVM methods were combined for effective individual forecasting. The ARIMA model was adopted to forecast linear components, and the LSSVM method was used to forecast nonlinear components. Finally, individual predictions were assembled via a simple addition (ADD) approach to obtain the final prediction results. The proposed method is described in Fig.1.

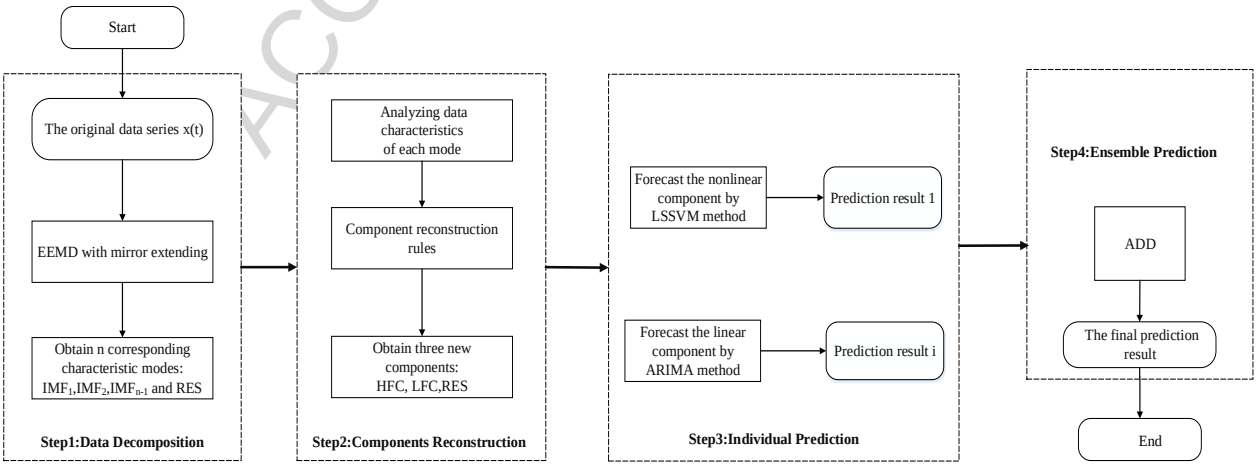

Fig. 1. Framework of the EEMD-LSSVM-ARIMA model

3.2. Data decomposition

The EEMD method was used to decompose the original dataset into a series of IMFs and one RES. The EEMD method is a noise-assisted data analysis method that is improved by empirical mode decomposition (EMD).

3.2.1. EMD method

The EMD method can handle non-linear, non-stationary time series (Huang et al., 1998) and has several significant advantages. The decomposition it performs is self-adaptive, uses orthogonal approximation, is highly efficient, and continuous. The decomposition results are also very conclusive. Local scale features from the original time series can be preserved, and the method can identify hidden patterns in the original data according to the dataset itself. The EMD is used to decompose complex time series and extract the characteristics of different time scales; several IMFs and one RES are obtained in the order of high to low frequency. The algorithm is as described below.

Step 1: Determine all of the maximum and minimum values from the original time series $x(t)$, then use the cubic spline function to generate upper and lower envelopes, $x_{\max}(t)$ and $x_{\min}(t)$.

Step 2: Calculate the point-by-point mean $m(t)$ from the upper envelopes $x_{\max}(t)$ and lower envelopes $x_{\min}(t)$:

$$m(t) = \frac{x_{\max}(t) + x_{\min}(t)}{2}$$

Step 3: Separate the first IMF. The original time series $x(t)$ minus the mean $m(t)$ is $d(t)$ (i.e., $d(t) = x(t) - m(t)$). Then check whether $d(t)$ meets the following two IMF conditions:

(1) In the whole of the data series, the extremum number (sum of the maximum and minimum) is equal to the number of zero crossings, or differs by one at most.

(2) At any point, the arithmetic mean value of the upper and lower envelopes is zero.

If $d(t)$ meets the conditions, replace $x(t)$ with residual $r(t) = x(t) - d(t)$; if $d(t)$ does not meet the conditions, replace $x(t)$ with $d(t)$ and repeat the above steps until the condition is met.

Step 4: Separate other IMFs and RES. Repeat steps 1 to 3 for $x(t)$ to find other IMFs and residuals until the end criterion is satisfied.

The original time series can be expressed by the sum of IMFs and one RES:

$$x(t) = \sum_{i=1}^{n} d_i(t) + r(t)$$

where $d_i(t)$ is the i^{th} IMF, $r(t)$ is the residual, and n is the number of IMFs. The total number of IMFs is limited to $Log_2 N$, where N is the length of the original time series.

3.2.2. EEMD method

For the decomposition of non-stationary, non-linear data, the EMD method has advantages, but it is not without flaws. There are two obvious drawbacks of the EMD method: one concerns the end effect and the other concerns mode mixing. In the EMD process, cubic spline interpolation is introduced to create the upper and lower envelopes, which displays worse behavior at the two edges. Moreover, there may be false IMF components, which will gradually contaminate the entire dataset, resulting in distorted decomposition results (i.e., the end effect; Wu and Huang, 2009). To overcome these problems, many improved techniques have emerged, such as mirror extending, polynomial curve fitting, extremum extending, SVR extending, and neural networks extending (Cao and Zhang, 2013； Xiong et al., 2014). Each method has its advantages, but the end effect cannot be completely eliminated. We adopt mirror extending to reduce the end effect. The mode mixing problem, whereby time series with the same frequency are separated into different IMFs or time series with different frequencies are combined into an IMF, causes some of the IMFs to be illogical and distorts the findings (Lei et al., 2009). To overcome this problem, the EEMD method was modified by Wu and Huang (2009). The EMD is still the core of the method, but the white noise sequence is introduced before using the EMD for auxiliary decomposition. The final decomposition result is the integrated averages of the auxiliary decomposition results. The algorithm is as follows:

Step 1: Add white noise $w(t)$ to the original data series $x(t)$ to obtain a new data series $x_i(t)$; thus

$x_i(t) = x(t) + w(t), (i = 1 \cdots M)$.

Step 2: Decompose the new data series $x_i(t)$ into several IMFs using EMD;

Step 3: Repeat steps 1 and 2 M times, each time adding random white noises;

Step 4: Calculate the mean value of the corresponding IMFs obtained from each decomposition. The integrated average value is the final decomposition result.

By adding the white noise sequence, the likelihood of mode mixing is reduced, and the impact of white noise can be offset by the integrated average, which represents a considerable advance over the EMD method. The effect of the added white noise sequence can be controlled by the final standard deviation of error (ε_M):

$$\varepsilon_M = \frac{\varepsilon}{\sqrt{M}}$$

where M is the ensemble's numbers, ε is the standard deviation of the added white noise, and ε_M is the final standard deviation of error. In practice, M is usually set to 100 or 50 and ε to 0.1 or 0.2 (Wu and Huang, 2009).

3.3. Component reconstruction

Not all decomposition modes contain the changing characteristics of the original data series. To provide the modes with clearer economic meaning and reduce the computational cost of individual predictions, the modes are further reconstructed into meaningful components. The reconstruction process is divided into two stages. The first is a statistical analysis of all the decomposed modes in order to identify any reconstruction implications (Zhang et al., 2014). Second, a t-test is conducted to obtain the reconstruction rules (Zhang et al., 2008).

3.3.1. Data characteristics analysis

To identify the key hidden data characteristics, all of the decomposed modes are statistically analyzed. There are four main analysis indicators: mean period, correlation coefficient, variance, and variance contribution rate. The mean period is the value obtained by dividing the length of the data series by the number of peaks of each mode.

The Pearson product-moment and Kendall rank correlation coefficients are used to measure the degree of correlation between the modes and the original data series from different perspectives. Variance represents the fluctuation level of each mode. The contribution of each mode to the total volatility of the observed data series is explained by the variance percentage of the observed data series. The sum of the variance of each mode is not always equal to the observed variance due to the nonlinearity of the original data series, a combination of rounding errors, and the introduction of variance through the treatment of the cubic spline end conditions (Peel et al., 2005). A variance percentage of $(\sum IMF + RES)$ is thus adopted as a supplement. The indicators and reconstruction implications are shown in Table 2.

Table 2

Data characteristics analysis

Indicators	Data characteristics	Reconstruction implications
Mean period	The shorter the mean period, the higher the fluctuation frequency, the greater the difficultly in identifying characteristics and the greater the difficulty in forecasting results.	Combine this with other similar modes to avoid amplifying prediction errors and reduce computational cost.
Pearson correlation and Kendal correlation	The greater the correlation, the more characteristics are contained, and the easier it is to forecast results.	Combine the modes with close correlations to identify frequent economic connotations.
Variance	The greater the variance, the more irregular the volatility, the greater the difficulty in identifying characteristics, and the more difficult it is to forecast.	Combine modes with high variance to avoid amplifying prediction errors and combine the modes with small variance to reduce computational costs.
Variance as % of observed	The higher the variance contribution rate, the more characteristics are contained, and the easier it is to forecast.	Combine the modes with close variance contributions to reduce computational costs.
Variance as % of $(\sum IMF + RES)$	The higher the variance contribution rate, the more characteristics are contained, and the easier it is to forecast.	Combine the modes with close variance percentages to reduce computational costs.

3.3.2. Reconstruction rules

Based on the reconstruction implications, the t-test of the mean of the sum of each mode is introduced to generate reconstruction rules. Each IMF has a different time scale, so HFC and LFC can be obtained by

combining IMFs according to frequency, from high to low, and the residual is treated separately. The process is as follows.

Step 1: Calculate the mean of the sum of each IMF d_1 to d_i, redefine the indicator i (i.e., IMF_1) as indicator 1 by analogy, redefine the sum IMF_1 and IMF_2 as indicator 2, and redefine the sum of IMF_1 to IMF_i as indicator i;

Step 2: Assess which means significantly differ from zero and identify indicator i;

Step 3: According to the result of step 2, reconstruct the IMFs. IMF_1 to IMF_{i-1} are defined as high-frequency components (HFCs), IMF_i to IMF_n are defined as low-frequency components (LFCs), and the residual is the trend component (RES).

Through the reconstruction，HFC $h(t)$, LFC $l(t)$, and RES $r(t)$ are obtained with more significant structural features and economic meaning than IMFs; therefore, the original data series $x(t)$ can be expressed as follows:

$$x(t) = h(t) + l(t) + r(t)$$

3.4. Individual prediction

Traditional forecasting methods do not take randomness, periodicity, or the trends of the original time series into account, leading to omissions and losses of information and affecting the accuracy of the prediction result. To overcome these drawbacks, we use individual prediction for the reconstructed components. The ARIMA model is a class of linear model and can be used to capture the linear patterns existing in the data series. The LSSVM has excellent nonlinear modeling capability and is used to account for the nonlinear patterns hidden in the data series.

3.4.1. LSSVM model

The LSSVM model is a modification of the standard SVM model. The SVM is based on the principle of structural risk minimization, which emphasizes the minimization of the upper promotion error boundary (Zhu

and Wei, 2013). The SVM has several unavoidable drawbacks. For instance, the complexity of its algorithm depends on the sample data size: the greater the sample data, the slower the calculation and the more time required for training. The LSSVM replaces the quadratic programming optimization of the SVM by solving a set of linear equations, which reduces computational complexity and accelerates the solution time (Zhang et al., 2015). The LSSVM regression algorithm is as follows:

Step 1: For a given set of training samples $(x_i, y_i), x_i \in R^n, y_i \in R, (i = 1, 2, ...l)$, nonlinear mapping ϕ is used to map the dataset from the input to high-dimensional feature spaces, so that the problem of nonlinear fitting in the input space becomes a problem of linear fitting in the high-dimensional feature space. The linear regression function of the high-dimensional feature space can be expressed as follows:

$$y_i = w^T \phi(x_i) + b$$

where w and $\phi(x_i)$ are n dimensional vectors, and b is the bias term.

Step 2: According to structural risk minimization principles and considering functional complexity and fitting errors, the regression problem can be expressed as a constraint optimization problem, as follows:

$$\min_{w,b,e} J(w,e) = \frac{1}{2} w^T w + \frac{\gamma}{2} \sum_{i=1}^{l} e_i^2;$$
$$s.t. y_i = w^T \phi(x_i) + b + e_i, (i = 1, 2, ...l)$$

where e_i is the training error at time i, and γ represents the regularization constant.

Step 3: To solve the above optimization problem, the constraint optimization problem is transformed into an unconstrained optimization problem. Based on the Lagrangian function and Kuhn–Tucker conditions for optimization, the original problem can be described as follows:

$$y_i(x_i) = \sum_{i=1}^{l} \partial_i K(x, x_i) + b$$

where $K(x, x_i)$ is the Kernel function, which has a variety options. Typical kernel functions include polynomial, radial basis (RBF), and Sigmoid kernel functions. An RBF kernel function can yield an accurate prediction, and is thus commonly used. We use an LSSVM method with an RBF kernel function; thus,

$k(x, x_i) = \exp(-\| x - x_i \| / 2\sigma^2)$, where σ^2 is the squared bandwidth, optimized by an external optimization tool during the training process.

The selection of RBF kernel parameters σ^2 and panel factor γ is a critical question. Empirically, the grid search algorithm is a valid and common method used to obtain the optimal determination of (γ, σ^2). A detailed explanation of this optimization technique can be found in Wu et al. (2008) and is thus not repeated here.

3.4.2. ARIMA model

The ARIMA model stems from the auto-regressive (AR) and moving average (MA) models as well as a combination of the AR and MA (ARMA) models (Wang et al., 2015). The ARMA model can be expressed as follows:

$$x_t = \varphi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}$$

where x_t is the predicted or simulated value, p and q represent the values of the autoregressive and moving average terms, $\varphi_1, \varphi_2, \cdots, \varphi_p$ are coefficients associated with each previous original value, $\theta_1, \theta_2, \cdots \theta_q$ are coefficients associated with previous white noise values, and a_t (hypothetical white noise) should be independently and identically distributed, with zero mean and variance σ^2.

The ARMA model is used for stationary time series, expressed as ARMA(p, q). If the time series is non-stationary, it should be transformed into a stationary time series by the $d'th$ difference process, as follows:

$$\varphi(B)\nabla^d x_t = \theta(B) a_t$$

where B is the backwards operator, and $\nabla^d x_t$ is the stationary series after the difference process.

We therefore expressed it as ARIMA (p, d, q), where p represents the number of auto-regressive terms, d represents the difference, and q represents the number of moving average terms. The ARIMA model is built in four main steps:

Step 1: Check the stability of the data series. Test the smoothness of the time data series, especially to

check whether it contains the unit root and the number of unit roots. If there is a unit root, the data are in a non-stationary sequence, which requires differential stabilization.

Step 2： Determine the structure of the ARIMA model. Some forms of the model can be initially identified using statistical methods such as autocorrelation (AC) and partial autocorrelation (PAC) coefficients that can represent the characteristics of the data series. Optimal parameters are determined according to the order of the Akaike Information Criterion (AIC) or Schwarz Criterion (SC), and a corresponding optimal model is selected from among those available;

Step 3: Test the performance of the ARIMA model, the significance of the model parameters, the validity of the model itself, and whether the residual sequence is "white noise." If the model passes these tests, the model setting is correct; otherwise, the model must be redefined and retested until the correct model form is obtained;

Step 4: Apply the ARIMA model to predict and compare the forecast and actual values.

3.5. Ensemble prediction

The original time series are decomposed into different modes and reconstructed into meaningful components from high to low frequency. The sum of these components is equal to the actual value of the original time series. Thus, for ensemble prediction, a simple but effective approach that sums the individual prediction results of each component is employed. The results can be expressed as follows:

$$\hat{x}(t) = \hat{h}(t) + \hat{l}(t) + \hat{r}(t)$$

where $\hat{x}(t)$ is the final prediction result, $\hat{h}(t)$ is the HFC forecast value, $\hat{l}(t)$ is the LFC forecast value, and $\hat{r}(t)$ is the RES forecast value.

3.6. Forecasting evaluation criteria

To measure the prediction performance, several criteria are selected. Mean absolute percent error (*MAPE*), root mean square error (*RMSE*) are adopted to evaluate the prediction level. A useful statistic (D_{stat}) is also adopted

to evaluate the directional forecasting. The three criteria are defined as follows:

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{x(t)-\hat{x}(t)}{x(t)}\right|,$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}[\hat{x}(t)-x(t)]^2},$$

$$D_{stat} = \frac{1}{N}\sum_{t=1}^{N}a_t \times 100\%,$$

where $x(t)$ and $\hat{x}(t)$,($t = 1,2,...,N$) are, respectively, the real and forecasting values at time t, and N is the data size of the testing set. Here, $a_t = 1$ if $[x(t+1)-x(t)][\hat{x}(t+1)-\hat{x}(t)] \geqslant 0$; otherwise, $a_t = 0$. Obviously, the smaller the $MAPE$ and $RMSE$ and the higher the D_{stat}, the higher the prediction accuracy, and the better the forecast model's performance.

To confirm the superiority of the proposed hybrid method, other widely used prediction methods are examined for comparison. The ARIMA and LSSVM models are employed as single benchmarks, and the widely used hybrid EEMD-LSSVM and EEMD-ARIMA models are considered.

4. Data and parameters

4.1. Data source

The amount of coal overcapacity is the difference between coal capacity and actual coal output, which can be calculated as follows:

$$OC = C - P = P / CU - P \qquad (1)$$

where OC is the coal overcapacity scale, C is the coal production capacity, P is the actual coal output, and CU is the coal capacity utilization. The data used for coal output were collected from a survey conducted by the China National Bureau of Statistics. Here, actual coal output covers the period from January 1989 to December 2016 in monthly units. Coal capacity utilization was obtained by calculating the mean value of the three different meanings of capacity utilization (i.e., ECU, TCU, and PCU). The data used for ECU, TCU, and PCU covering 1989 to 2013 are obtained from existing studies (Zhao and Huang, 2014; Feng et al.,

2015; Zhang, 2014). The data for *ECU* , *TCU* , and *PCU* covering 2014 to 2016 are continuously calculated based on the three above mentioned studies, and the relative data used for calculation are taken from the *China Energy Statistical Yearbook* and the *China Coal Industry Statistical Yearbook* (2014–2016).

We considered three different types of coal capacity usage types: economic capacity utilization (*ECU*), calculated over the cost function; technical capacity utilization (*TCU*), calculated via data envelopment analysis; and physical capacity utilization (*PCU*), calculated via the production function. Comparing these three different coal capacity uses, calculated using three different methods, generated closed values. To ensure data rationality, we took the arithmetic mean of *ECU* , *TCU* , and *PCU* as the value of *CU* . The results are shown in Table 3.

Table 3

The value of coal capacity utilization

Year	*ECU*	*TCU*	*PCU*	*CU*=(*ECU*+*TCU*+*PCU*)/3
1989	--	66.78%	69.27%	68.03%
1990	--	69.95%	70.36%	70.16%
1991	--	70.59%	65.87%	68.23%
1992	--	65.06%	64.65%	64.86%
1993	--	66.28%	62.08%	64.18%
1994	71.42%	68.56%	64.77%	68.25%
1995	71.19%	69.03%	67.25%	69.16%
1996	73.69%	70.65%	67.58%	70.64%
1997	71.57%	65.99%	65.24%	67.60%
1998	65.75%	66.92%	64.14%	65.60%
1999	67.41%	58.31%	53.70%	59.81%
2000	67.78%	71.06%	67.92%	68.92%
2001	76.38%	82.25%	64.39%	74.34%
2002	83.59%	83.90%	73.13%	80.21%
2003	83.10%	84.39%	86.02%	84.50%
2004	87.56%	85.89%	93.23%	88.89%
2005	76.37%	87.02%	95.28%	86.22%
2006	78.58%	89.86%	92.67%	87.04%
2007	81.78%	93.64%	91.21%	88.88%
2008	77.36%	88.98%	88.20%	84.85%
2009	91.50%	85.42%	86.59%	87.84%
2010	86.44%	86.73%	88.26%	87.14%
2011	82.47%	88.56%	91.16%	87.40%

2012	68.25%	69.89%	69.95%	69.36%
2013	78.03%	70.25%	71.59%	73.29%
2014	76.01%	72.17%	73.98%	74.05%
2015	72.02%	69.87%	71.05%	70.98%
2016	65.29%	61.44	69.38%	65.37%

Notes: The data for ECU are taken from Zhao and Huang (2014), the data for TCU are taken from Feng et al. (2015), and the data for PCU are taken from Zhang (2014).

The sample data (i.e., on the amount of coal overcapacity) are obtained based on the collated data described above and calculations for January 1989 to December 2016, a total of 336 monthly data points. According to the coal overcapacity series（see Fig. 2）, the maximum amount of coal overcapacity is 1.656 billion tons and occurs in December 2016; the minimum amount of coal overcapacity is 1.208 billion tons and occurs in January 2004; and the average and the median amounts of coal overcapacity are 0.532 billion tons and 0.415 billion thousand tons, respectively. The sample dataset is divided into two parts: the training set and testing set. Data from January 1989 to December 2015 are used as the training set for model training, while data from January 2016 to December 2016 are used as the testing set to evaluate model performance.

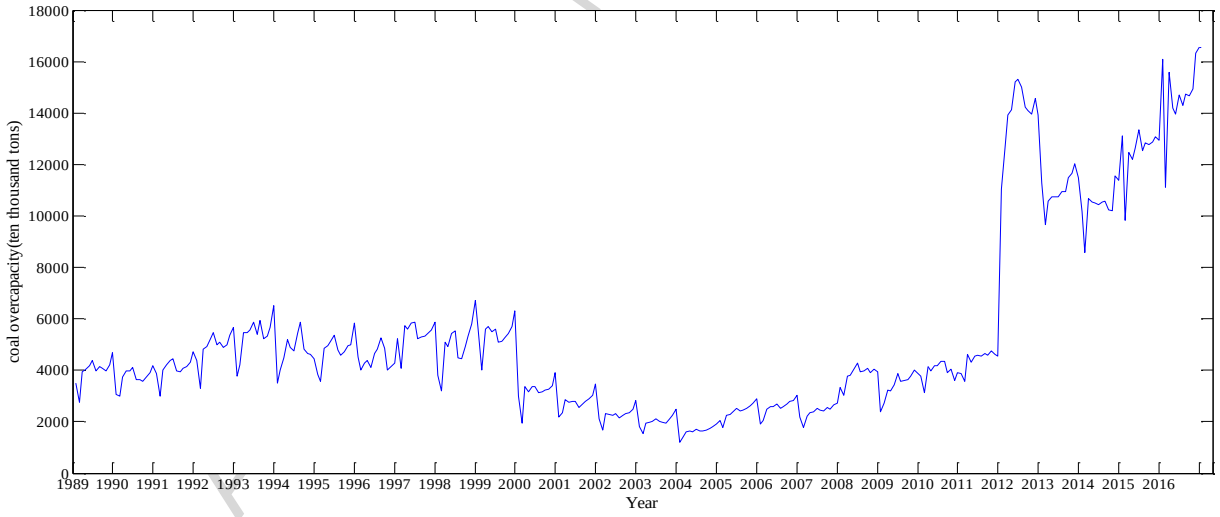

Fig. 2. Coal overcapacity data series for China from Jan. 1989 to Dec. 2016.

4.2 Parameter determination

Two parameters need to be determined in the EEMD process: the ensemble number M and the standard deviation of the added noise ε, which directly influences decomposition performance. Wu and Huang (2009) suggest that the ensemble number is often set to 50 or 100 while the standard deviation of the added noise is

often set to 0.1 or 0.2. We set the ensemble number and the added white noise amplitude (M, ε) to (100, 0.1), (50, 0.1), (50, 0.2), and (100，0.2) respectively. The decomposition results show that, when the ensemble number is 100 and the stand deviation of the noise is 0.2, the EEMD model performed best. Thus，M is set to 100 and ε is set to 0.2.

When the four benchmark models are introduced, the four sample datasets (Observed, HFC, LFC, and RES) are predicted using the LSSVM and ARIMA methods respectively. When using the LSSVM method of the RBF kernel function，the panel factor γ and kernel parameters σ^2 need to be established. Since the two parameters greatly affect prediction accuracy, the grid search technique is selected as an optimization tool for the parameter values. When using the ARIMA model, three parameters should be evaluated (i.e., ARIMA(p,d,q)). The d value can be determined by the difference order, the possible values of p and q can be selected by the auto-correlation coefficient (ACF) and partial correlation coefficient (PACF), and the optimal parameters can be determined by the lowest AIC and SC. The parameter determination process is detailed in the appendix, part B. Table 4 shows the optimal parameters of the LSSVM and ARIMA models.

Table 4

Optimal parameters of LSSVM and ARIMA models

Parameters of sample data set	LSSVM		ARIMA		
	γ	σ^2	p	d	q
Observed	24593.92	321.35	1	1	1
HFC	1870.48	95.76	2	0	1
LFC	456.37	36.08	3	1	1
RES	160.03	16.79	1	0	1

5. Results and discussion

5.1. Data decomposition

With EEMD selected as the multiscale decomposition tool for the coal overcapacity series, the model is established using the EEMD tool box of the Matlab software package. To limit the end effect of the EEMD, the

mirror extending method is used.[3] Through the EEMD with a mirror extending method, coal overcapacity data series is decomposed using MATLAB R2014a programming. The decomposition process produced seven IMFs and one RES, all of which are listed in the order in which they are sifted in Fig. 3. Each IMF presents a different fluctuation structure. With frequencies varying from high to low, the amplitudes of the decomposed modes are increasing.

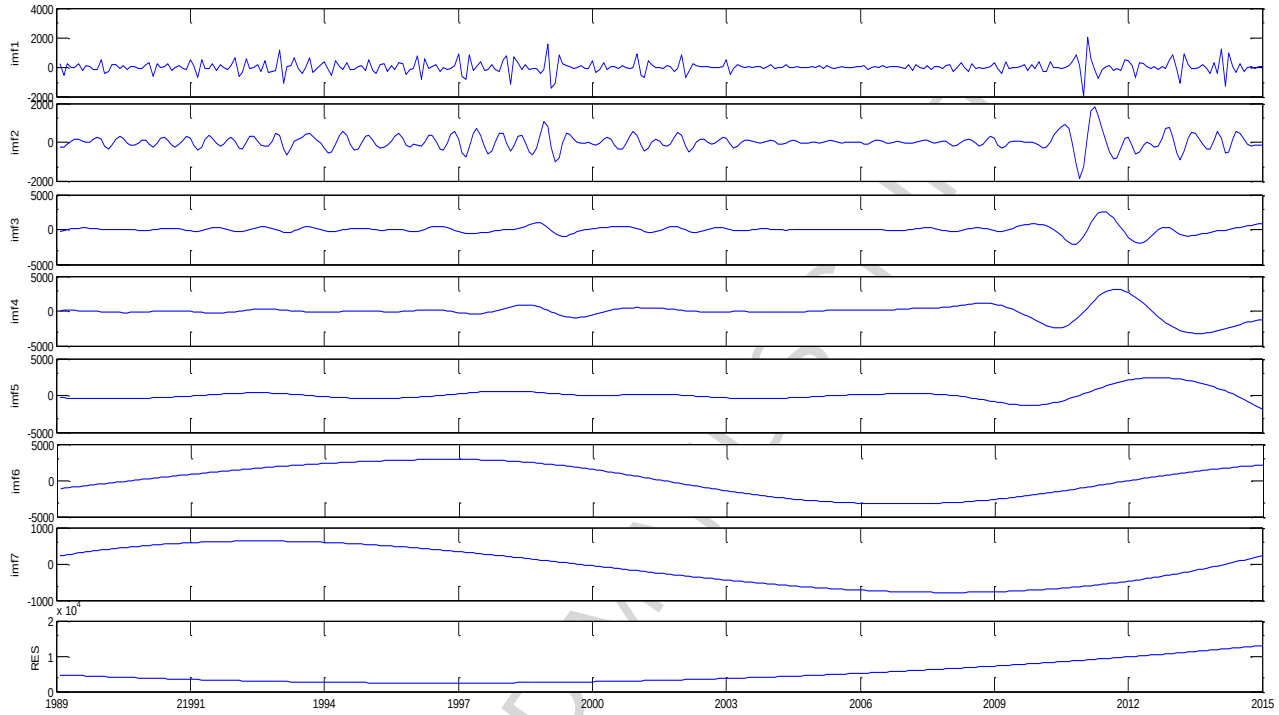

Fig. 3. Decomposition results of the observed data series via the EEMD with mirror extending method

To explore the variability and fluctuation characteristics of the coal overcapacity series from a new perspective, statistical analyses of the IMFs and RES are conducted. The results are as follows (see Table 5):

- The mean period lengths of IMF1, IMF2, and IMF3 are three, six, and nineteen months, respectively. The Pearson coefficients are 0.142, 0.159, and 0.166, respectively, and the Kendall coefficients are 0.059, 0.099, and 0.105, respectively, indicating that the correlations of IMF1, IMF2, and IMF3 with the observed series are very low. The contribution rate of the variance is about 2%, indicating that IMF1, IMF2, and IMF3 explain less of the fluctuation in the coal overcapacity data series than the other modes do. These three

[3] At the same time, we use another common support vector machine regression extending method for comparison. The results (see appendix, part A) show that, for the coal overcapacity series of this study, the mirror extension method is better at inhibiting the end effect.

modes all possess the characteristics of a short mean period and frequent fluctuations, suggesting that IMF1, IMF2, and IMF3 contain more uncertain and random points. These three modes thus have a low correlation with coal overcapacity and offer a weak explanation for the volatility of the coal overcapacity series.

- The mean period lengths of IMF4, IMF5, IMF6, and IMF7 are more than three years, thus longer than for the first three. The contribution rates of these modes to the variance are also greater, suggesting that they contain some significant aspects of the volatility of the coal overcapacity data series.
- Both the Pearson and Kendall coefficients, between the RES and coal overcapacity data series, reach a peak that is greater than 0.594 and 0.664, respectively, and the RES variances account for more than 70% of the total variability of the observed series, indicating that the residual mode is dominant and can describe the main fluctuation patterns of the coal overcapacity data series.

Table 5

Results of IMFs and RES statistical analyses

	Mean period (month）	Pearson correlation	Kendall correlation	Variance	Variance as % of observed	Variance as % of $(\sum IMF + RES)$
Observed				1.16E+07		
IMF1	3	0.142*	0.059	1.56E+05	1.34%	1.02%
IMF2	6	0.159*	0.099**	1.36E+05	1.16%	0.88%
IMF3	19	0.166*	0.105**	3.07E+05	2.64%	2.01%
IMF4	40	-0.065**	0.023**	1.05E+06	9.01%	6.85%
IMF5	81	0.597**	0.331**	5.69E+05	4.89%	3.71%
IMF6	162	0.475**	0.454**	4.18E+06	35.88%	27.26%
IMF7	162	0.203*	0.283**	2.59E+05	2.22%	1.69%
RES	324	0.594**	0.664	8.66E+06	74.46%	56.57%
SUM				1.53E+07	131.61%	100.00%

Notes: * and ** indicate statistical significance at the 5 and 1% percent levels, respectively (2-tailed).

5.2. Component reconstruction and discussion

5.2.1. Component reconstruction

For each component, the mean fluctuation period, the degree of correlation with the observed data series, and

the interpretation of the volatility of the original series differ. Some of the seven IMFs have analogous structural variations, indicating that they have similar economic implications. They are therefore further reconstructed into meaningful components, with a more fundamental economic meaning and new features according to their data characteristics. The residual is treated separately, which can explain the volatility of the original series more accurately and thus identify the causative factors of Chinese coal overcapacity and its evolution trends. The mean of the reconstruction as a function of the IMF index i is shown in Fig. 5. The mean values depart significantly from zero at IMF4; consequently, IMF1, IMF2, and IMF3 are reconstructed into high-frequency components (HFC); IMF4, IMF5, IMF6, and IMF7 are reconstructed into low-frequency components (LFC); and RES is treated as a trend component. Fig. 5 shows the three components.

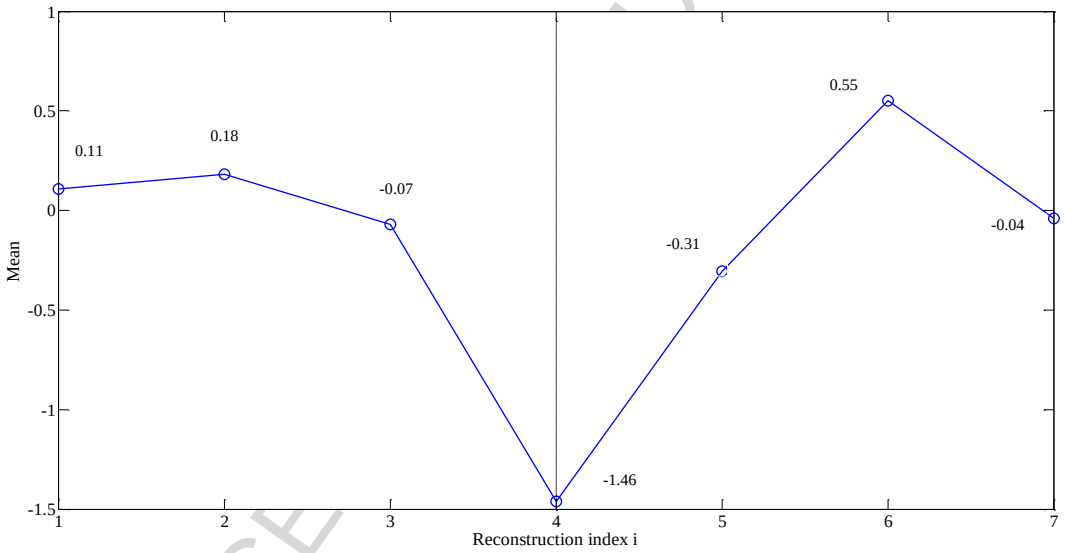

Fig. 4. Mean of fine-to-coarse reconstruction as a function of index i. The vertical dashed line at i = 4 indicates that the mean significantly departs from zero ($p < 0.01$).

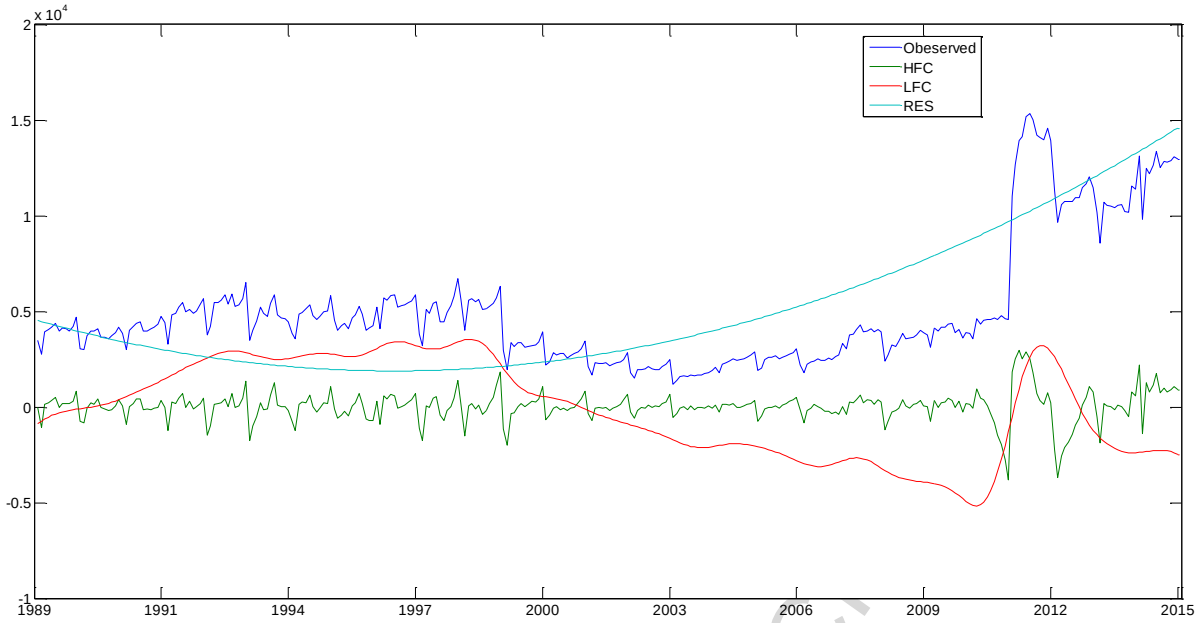

Fig. 5. Three components of the overcapacity series from Jan. 1989 to Dec. 2015.

The frequencies and amplitudes of HFC, LFC, and RES differ, and each has distinct characteristics. Table 6 presents the results of the statistical analysis. The Pearson and Kendall coefficients of HFC are 0.239 and 0.225, and the Pearson and Kendall coefficients of LFC are 0.528 and 0.529, respectively. Following reconstruction, the degree of correlation is improved over the correlations prior to reconstruction. The percentage of variance of HFC and LFC is about 4% and 44%, respectively, indicating that LFC's interpretation of the volatility of the observed data series is greater than that of HFC. Meanwhile, RES has the highest correlation with the observed data series and accounts for more than 70% of the volatility, indicating that most of the volatility of the original data series can be explained by RES. Regarding the influence of the three components on the volatility of the observed data series, the most important is RES, followed by LFC, and HFC's contribution is the lowest.

Table 6

Results of statistical analysis of HFC, LFC, and RES

	Mean period (month)	Pearson correlation	Kendall correlation	Variance	Variance as % of observed	Variance as % of $(\sum IMF + RES)$
Observed				1.16E+07		
HFC	6	0.239**	0.225**	7.45E+05	6.40%	4.38%

LFC	48	0.528*	0.529**	7.62E+06	65.46%	44.74%
RES	324	0.594**	0.664	8.66E+06	74.46%	50.89%
SUM				1.70E+07	146.32%	100%

Notes: * and ** indicate statistical significance at the 5 and 1% percent levels, respectively (2-tailed)

5.2.2. Component discussion

The above analyses show that each component possesses distinct characteristics and has a definite implication that can explain the volatility of coal overcapacity. According to Lian and Wang (2012), Zhang et al. (2017), and Fan et al. (2015), the main causes of overcapacity in China's coal industry include insufficient demand, market failure, and institutional distortion. However, these studies did not scientifically determine the extent to which these factors contribute to Chinese coal overcapacity. To address this question, we conducted quantitative and multiscale analyses of the fluctuation in Chinese coal overcapacity based on the results of the statistical analysis of the reconstructed components. The findings are as follows:

(1) HFC fluctuates around the zero-mean line with a small amplitude and contributes to 6.40% of the variability of the coal overcapacity data series. The HFC trend is the same in the observed data series at most data points, indicating that, during the entire period of the original time series, HFC has very little influence on the observed series but can influence a certain part of the fluctuation trend in the coal overcapacity series in a short term period. HFC can fully describe the impact of insufficient demand on excess coal capacity because the coal industry, a basic national industry, is more sensitive to market demand changes (Wang et al., 2017) than general processing industries are. There is a significantly negative correlation between HFC and China's GDP growth rate (see Fig. 6): when GDP growth slowed, the scale of coal overcapacity increased; when GDP growth recovered, coal overcapacity decreased. For example, amid the 2008 global financial crisis, global economic growth slowed due to a decline in energy demand, while China's national economic structure and growth pattern further exacerbated the imbalance of coal supply and demand. Thus, coal supply far exceeds demand, capacity utilization is inadequate, and the excess coal capacity exhibits a growth trend caused by declining market demand. Therefore, we may conclude that the small fluctuations in short-term coal overcapacity are

driven primarily by insufficient coal demand.

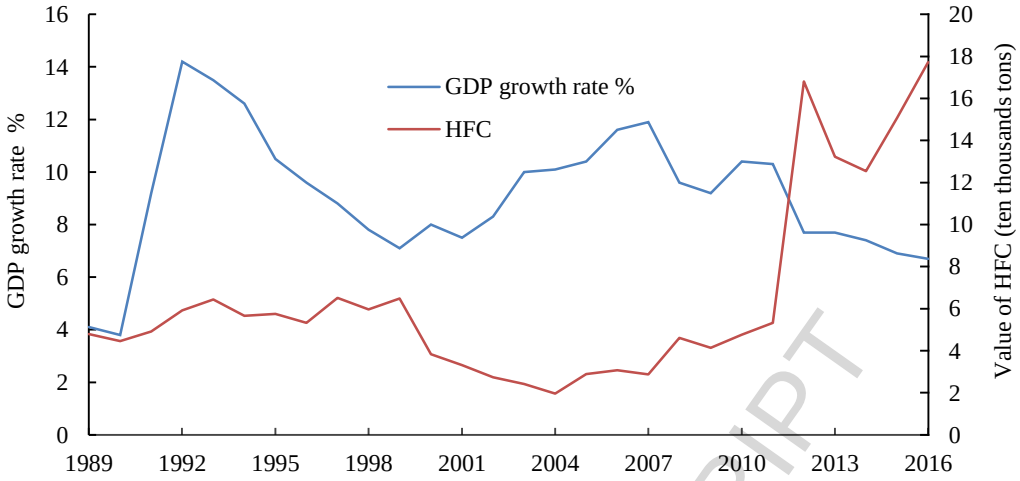

Fig. 6. GDP growth rate and HFC data series

(2) LFC accounts for 65.46% of the volatility and has a stronger correlation with the observed series, indicating that this component contains important features of the observed series. The amplitude of LFC is larger than that of HFC, but the curve is smoother, suggesting that the effect of this component on the observed series is more significant and its duration is longer. The fluctuation characteristics of LFC (see Fig. 5) show that there is an increasing trend from 1989 to 1999, a steady decrease from 2000 to 2011, and a rapid rise after 2012. Revisiting the history of investment in the coal industry, we found that the volatility of LFC is consistent with the investment wave phenomenon in the Chinese coal market. For example, China's economy entered a rapid growth period in 1989 (see Fig. 6), and a large number of enterprises and capital entered the coal industry, anticipating rapid growth in coal demand. However, due to China's reduced economic output, there is relatively little coal demand; thus, LFC is increasing from 1989. Additionally, during the 11th Five-year Plan period, fixed asset investment in the Chinese coal industry reached 1.25 trillion yuan, which was more than double the total investment during the 10 largest five-year plans. Despite the financial crisis and the declining coal demand in China, coal supply is still increasing, leading to rapid growth in coal overcapacity. LFC shows a sharply increasing trend in 2012. According to the wave phenomenon theory (Lin et al., 2010), therefore, LFC may quantitatively describe the impact of market failure on Chinese coal overcapacity.

(3) RES has the highest correlation with the observed series, and its contribution to the volatility of the observed series is 74.46%, which is significantly higher than those of HFC and LFC. Over the long term, the overall growth trajectory and development trend of Chinese coal overcapacity is consistent with that of RES, both of which have an increasing trend, indicating that the RES is a deterministic force for coal overcapacity evolution in the long run. As Huang et al. (1998) pointed out, the residual is often treated as the determining long-term factor. Accordingly, the RES, as the determinant of the volatility of the observed series, may be interpreted as the long-term impact of institutional distortions on Chinese coal overcapacity. Government intervention in China's coal industry is profound, as coal is a basic energy resource. This study postulates that the origin of China's overcapacity is the hysteresis of China's institutional mechanisms and subsequent distortions in resource allocation (Zhang et al., 2017; Fan et al., 2015; Yu and Zhang, 2014). From an endogenous perspective, the prolonged coal overcapacity in China was caused by non-market causes. In addition, focusing on the long cycle and continuing increasing trend of RES, we may conclude that the root cause of Chinese coal overcapacity may be China's inefficient and unscientific government mechanisms, which cannot be explained by common market economy theory.

5.3. Individual and ensemble prediction

5.3.1. Prediction performance comparisons

Forecasting strategies can be employed according to the multiscale analyses of each component. We use the proposed new hybrid method to forecast the coal overcapacity scale of 2016 and compare the result with the actual value to measure the prediction performance. The HFC and LFC are non-linear components, and the RES is an approximately linear component. For individual forecasting via the EEMD-LSSVM-ARIMA model, we adopt the LSSVM method to forecast the HFC and LFC, as it can remarkably capture the linearity, and adopt a class of model, the ARIMA model, to forecast the RES. We also consider the single models (ARIMA and LSSVM) and the hybrid models (EEMD-ARIMA and EEMD-LSSVM) as the comparative methods. The single

models directly forecast the observed series with no decomposition. The hybrid models forecast the reconstruction components (HFC, LFC, and RES) without identifying the linear and non-linear patterns. The parameters of these models are detailed in Section 4.2.

Figs. 7 and 8 show the out-of-sample forecasting results of the models. The novel hybrid EEMD-LSSVM-ARIMA model clearly has higher prediction accuracy. The comparative forecasting performance results are summarized in Table 7. The three selected evaluation criteria statistically prove that the EEMD-LSSVM-ARIMA method outperformed the other considered benchmark models. Among the five models, the proposed model not only has the highest forecasting accuracy (due to its lower MAPE and RMSE values) but also achieves the highest hit rates in directional prediction ability (due to its highest D_{stat} value). The LSSVM tool avoids over-fitting the HFC while fully fitting the LFC, and the ARIMA model captures the linear fluctuations of the RES better than the other methods do, thus improving prediction accuracy. Overall, the EEMD-LSSVM-ARIMA model proposed on the basis of the "decompose–reconstruction–ensemble" principle outperformed the single and hybrid forecasting models considered in this study.

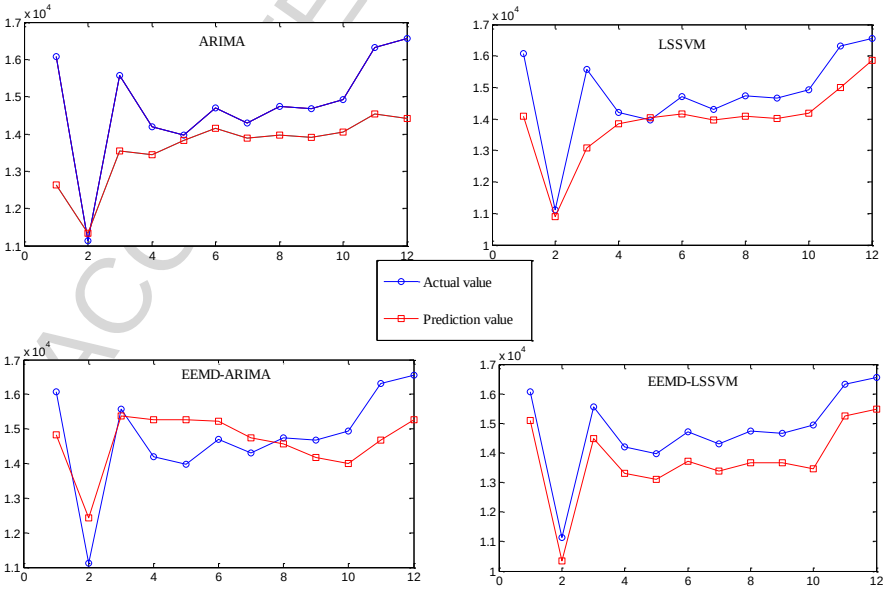

Fig. 7. Out-of-sample forecasting of four benchmark models.

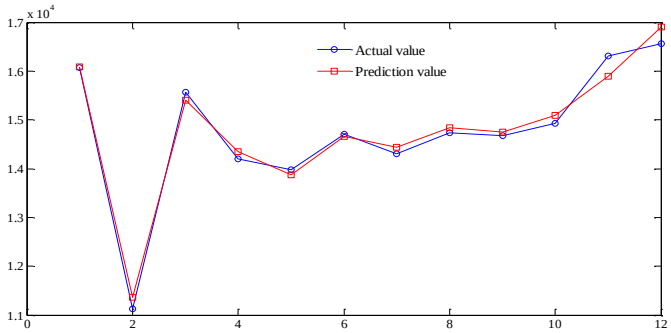

Fig. 8. Out-of-sample forecasting of EEMD-LSSVM-ARIMA model.

Table 7

Out-of-sample forecasting comparison of different models

Time	Actual value	ARIMA model		LSSVM model		EEMD-LSSVM model		EEMD-ARIMA model		EEMD-LSSVM-ARIMA model	
		Prediction value	Error (%)	Prediction value	Error (%)	Prediction value	Error (%)	Prediction value	Error (%)	Prediction value	Error (%)
1	16076.01	12633.98	21.41	14086.30	12.38	15084.37	6.17	14818.43	7.82	16088.86	0.08
2	11124.83	11341.57	1.95	10892.36	2.09	10348.90	6.97	12426.65	11.70	11362.49	2.14
3	15564.16	13552.2	12.93	13090.75	15.89	14484.41	6.94	15368.39	1.26	15407.12	1.01
4	14198.99	13441.49	5.33	13858.99	2.39	13314.42	6.23	15266.58	7.52	14346.26	1.04
5	13972.25	13825.01	1.05	14049.46	0.55	13108.03	6.19	15258.16	9.20	13869.05	0.74
6	14702.78	14152.54	3.74	14156.77	3.71	13713.69	6.73	15212.41	3.47	14663.90	0.26
7	14303.88	13884.25	2.93	13970.55	2.33	13384.09	6.43	14735.05	3.01	14443.52	0.98
8	14731.87	13964.52	5.21	14084.93	4.39	13670.92	7.20	14569.31	1.10	14839.60	0.73
9	14672.27	13910.64	5.19	14027.23	4.40	13671.44	6.82	14166.32	3.45	14749.39	0.53
10	14931.21	14044.37	5.94	14177.94	5.04	13449.67	9.92	13998.22	6.25	15096.94	1.11
11	16316.79	14542.47	10.87	15009.42	8.01	15238.02	6.61	14680.65	10.03	15894.61	2.59
12	16558.25	14407.53	12.99	15853.182	4.26	15482.49	6.50	15767.72	7.79	16908.44	2.11
MAPE (%)			7.46		5.45		6.89		6.05		1.11
RMSE			1492.93		1090.49		1030.70		1002.03		200.11
D_{stat}			0.83		0.83		0.83		0.75		1

Notes: The unit is ten thousand tons; error is an absolute value.

5.3.2. Forecasting results and discussion

As the EEMD-LSSVM-ARIMA model we propose has a high forecast accuracy, it can be used to predict the scale and development trends of the overcapacity in China's coal industry from 2017 to 2019, which can help guide national decision-makers, enterprises, and investors. The forecast results show that coal overcapacity in

China will continue to be very serious. Table 8 presents the specific forecast value and coal overcapacity scales for 2017, 2018, and 2019, which will be about 1.819, 1.721, and 1.762 billion tons, respectively. Fig. 9 shows the trend forecasts for each component and the original series. From 2017 to 2019, coal overcapacity will fluctuate within a certain range; in the short term, no significant increases or decreases similar to those in 2012 will occur. These results indicate that overcapacity will reach a relatively stable stage but will remain high.

Several findings can be drawn from the individual forecasts of three sub-series (i.e., HFC, LFC, and RES) from 2017 to 2019 (see Fig. 9), First, from the HFC, the fluctuation frequency will become lower than in the 2014 to 2016 period, but the amplitude will increase, indicating that the impact of declining demand on Chinese coal overcapacity caused by economic recession will slow but will remain noticeable. Second, the LFC will experience negative growth, indicating that large-scale policy regulation and industrial adjustment will limit the growth of coal overcapacity, and the effects of de-capacity and other coal industry control policies via China's supply-side reforms will be achieved. Third, the forecasted trend for the RES is rising, indicating that coal overcapacity is deeply influenced by institutional distortions that have a continuous effect and thus suggesting the potential for a long-term trend of excess among coal enterprises. Overall, China's coal overcapacity cannot be eliminated quickly but will require a long-term solution.

Our forecast results show that, with the vigorous implementation of de-capacity policies, China's coal overcapacity will not significantly increase or decrease between 2017 and 2019. However, due to the energy revolution and economic recession, the excess capacity problem in China's coal industry needs to be resolved. The supply and demand imbalance is serious, and de-capacity should be undertaken without delay. Under the guidance of national regulation, de-capacity goals were achieved easily in 2016; the targets were even surpassed. The current situation is not favorable, however. By 2020, coal demand will reach 4.03 billion tons, and the coal supply is projected to be at least 109 million tons greater than coal demand in China (Liu et al., 2017). Resolving the excess coal capacity problem, promoting a balance between coal supply and demand, and

accelerating structural coal supply reforms are key goals for the Chinese coal sector during the 13th Five-year Plan period.

Table 8

Ensemble prediction value of EEMD-LSSVM-ARIMA model from 2017 to 2019

Year / Month	2017	2018	2019
1	14966.08	14956.34	17172.91
2	15011.25	15041.90	16845.65
3	14929.98	15032.38	16345.32
4	15078.39	14555.23	16078.85
5	14979.10	14035.75	15624.29
6	15330.25	13555.43	15025.34
7	14935.33	12571.98	14690.58
8	14810.35	12377.21	13747.34
9	15829.09	14081.00	13345.85
10	15571.53	14662.63	12933.13
11	15456.56	14861.62	11390.98
12	14996.87	16399.83	13018.23
Sum	181894.78	172131.30	176218.47

Notes: The unit is ten thousand tons.

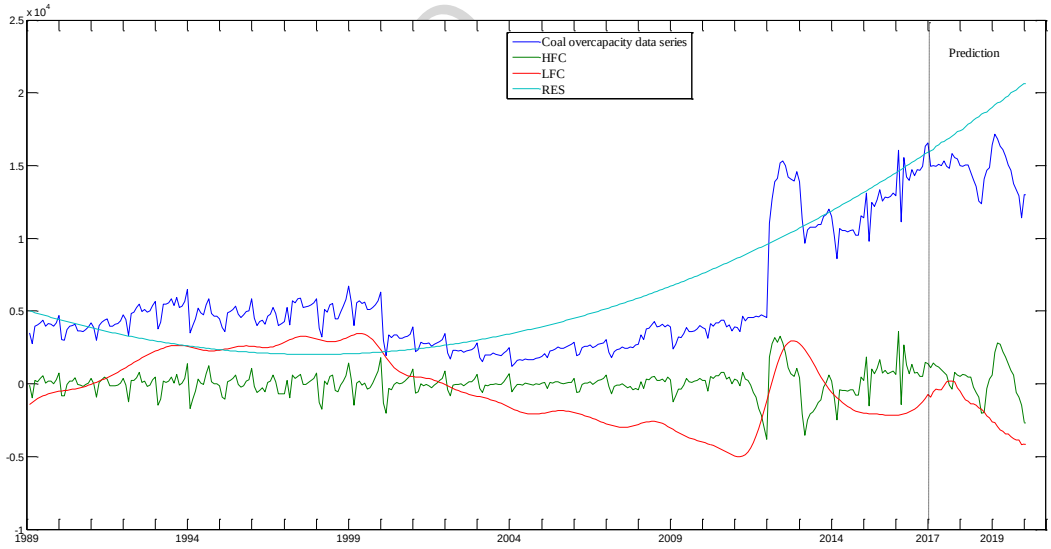

Fig. 9. Prediction results for coal overcapacity from 2017 to 2019

6. Conclusions and policy implications

6.1. Key conclusions

Considering the complexity and seriousness of the overcapacity in China's coal industry, we produced an

EEMD-LSSVM-ARIMA model to quantitatively analyze the causes and forecast the development trend of coal overcapacity in China. The results offer several important implications.

First, in the long term, institutional distortion is the most important cause of coal overcapacity in China, with an impact degree of 74.46%. This means that the root cause may be China's ineffective and unscientific government mechanisms, which cannot be explained by common market economy theory. Market failure is another important factor affecting coal overcapacity, with an influence degree of 65.46%, and has a long-lasting effect of about four years per cycle. Declining coal demand has little effect in the long term, with an influence degree of 6.40%, but it may exacerbate the coal overcapacity scale in the short term.

Second, the ensemble forecast results suggest that, between 2017 and 2019, Chinese coal overcapacity will total about 1.819, 1.721, and 1.762 billion tons, respectively, indicating that resolving excess capacity in China's coal industry will be difficult during the 13th Five-year Plan period. The individual forecasting evolution trends of the three components suggest that the effect of market failure on coal overcapacity will gradually weaken, while the negative effects of institutional distortions will be enhanced.

Finally, the EEMD-LSSVM-ARIMA model, based on the decomposition–reconstruction–prediction principle, has a promising level of forecasting power. The proposed EEMD-LSSVM-ARIMA model can capture the linear and nonlinear patterns in coal overcapacity data series, thus improving prediction accuracy, more effectively than other widely used single prediction models (e.g., ARIMA and LSSVM) or combined prediction models (e.g., EEMD-ARIMA and EEMD-LSSVM).

6.2. Policy implications

Based on the above conclusions, several policy recommendations for limiting coal overcapacity can be proposed to the Chinese government.

First, taking institutional regulation as a contact point, a strategy of institutional reform and innovative government intervention can solve this problem. Institutional distortions in China are manifested mainly in

over-investment caused by improper government regulation (Li and Zhang, 2015), redundant construction due to the GDP-led promotion system of governmental officials (Jiang et al., 2012), and the distortion of the factor price system and its determining mechanism under Chinese fiscal decentralization (Zheng, 2015). Therefore, solving the overcapacity problem must start by addressing those issues. In China's supply-side reform, coal capacity regulation policies should focus on and strengthen supply-side management and seek breakthroughs in capacity regulation from the supply side.

Second, our forecast indicates that, although the impact of market failure on overcapacity in the coal industry will decrease, it will remain high. Accordingly, to alleviate the wave phenomenon caused by informational imbalances, we suggest that the government establish a transparent monitoring system and perfect the disclosure mechanism for coal industry information. Specifically, the government should learn from foreign experience, refine its statistical system, and establish a statistical monitoring system for capacity utilization in the coal industry as soon as possible. The government has authority in many fields such as those concerning data and information and can implement wide-scale control policies. Thus, the government should establish a coal industry information release system and announce coal industry capacity scales and utilization rates, as well as other relevant information, on a regular basis and then guide market expectations, scientific investment, and production decisions in order to prevent harmful investment and expansion levels.

Third, as China's economy enters the "new normal" of high-speed growth, it will be difficult to reduce the excess coal capacity by relying on insufficient domestic demand because of the ongoing energy transformation and industrial restructuring. One novel way to solve China's low domestic coal demand is to seize the opportunity offered by China's economic new normal, seek projects driven by international cooperation, and make full use of the "One Belt One Road" strategic partnership. From the demand side, this could expand coal market demand in countries along the One Belt One Road, replace coal product output with coal capacity output, and digest the abundant coal capacity by breaking the limits of domestic coal market space. From the supply

side, this could expand direct investment and production capacity cooperation with countries along the One Belt One Road, enhance coal enterprises' willingness to participate in international cooperation areas, and realize capacity transfer through capacity cooperation, thus achieving complementary advantages and mutual benefits.

Acknowledgments

This research is supported by the Fundamental Research Funds for the Central Universities (No. 2017XKQY085).

References

Arfa C, Leleu H, Goaïed M, et al., 2017. Measuring the capacity utilization of public district hospitals in tunisia. Int. J. Health Policy Manag. 6(1), 9-18.

Aydin G., 2015. The Application of trend analysis for coal demand modeling. Energ. Source. PART B 10(2), 183-191.

Cao D, Zhang X., 2013. Research on the Comparison with Methods of Restraining Ending Effect of EMD and Its Application in Fault Diagnosis. J. Mecha. Trans.(3),83-87.

Dagdeviren H., 2016. Structural constraints and excess capacity: An international comparison of manufacturing firms. Dev. Policy Rev. 34(5), 623-641.

Davidson C, Deneckere R., 1990. Excess capacity and collusion. Int. Econ. Rev. (Philadelphia) 31(3), 521-54.

Ediger V S, Akar S., 2007. ARIMA forecasting of primary energy demand by fuel in Turkey. Energy Policy 35(3), 1701-1708.

Fan L K, Xiao-Ping L I, Ying S S., 2015. The realistic bases and formation mechanism of excess capacity based on incremental reform. Chin. Ind. Econ. 1, 19-31.

Feng D M, Wand S, ZHAI C X., 2015. Empirical research for coal industry capacity utilization rate estimation and influencing factors in China. Stat. Info. Forum. 183(12), 48-55

Feng Y Y, Chen S Q, Zhang L X., 2013. System dynamics modeling for urban energy consumption and CO2 emissions: a case study of Beijing, China. Ecol. Modell, 252, 44-52.

Greenblatt R E., 2017. Oscillatory dynamics of investment and capacity utilization. Physica A: Statistical Mechanics and its Applications 456, 486-493.

Höök M, Aleklett K., 2009. Historical trends in American coal production and a possible future outlook. Int. J. Coal. Geol. 78(3), 201-216.

Höök M, Zittel W, Schindler J, et al., 2010. Global coal production outlooks based on a logistic model. Fuel 89(11), 3546-3558.

Hu X, Zhao G., 2008. Forecasting model of coal demand based on matlab BP neural network. Chin. J. Manag. sci. 10(16), 521-525.

Hu Y C., 2017. Grey prediction with residual modification using functional-link net and its application to energy demand forecasting. Kybernetes 46(2), 349-363.

Huang N E, Shen Z, Long S R, et al., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. J. R. Soc. Interface 454(1971), 903-995.

Jain R K, Smith K M, Culligan P J, et al., 2014. Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy. Appl. Energy 123,168-178.

Jiang F T, Geng Q, Da-Guo L V, et al., 2012. Mechanism of excess capacity based on China's regional competition and market distortion. Chin. Ind. Econ. 12,18-31

Karagiannis R., 2015. A system-of-equations two-stage DEA approach for explaining capacity utilization and technical efficiency. Ann. Oper. Res. 227(1), 25-43.

Klein L R, Preston R S., 1967. Some new results in the measurement of capacity utilization. Am. Econ. Rev. 57(1), 34-58.

Lei Y, He Z, Zi Y., 2009. Application of the EEMD method to rotor fault diagnosis of rotating machinery. Mech. Syst. Signal Process. 23(4), 1327-1338.

Li L, Lei Y, Pan D., 2015. Economic and environmental evaluation of coal production in China and policy implications. Nat. Hazards 77(2), 1125-1141.

Li Y, Zhang X.,2015. The New Normal:the Logic and Perspective of Economic Development. Econ. Res. J. (5), 4-19.

Lian-Ji L I, Wang Y., 2012. A Study on Overcapacity Problem in China's Transition. On Econ. Prob.(12), 29-3.

Lin Y F, Wu H M, Xing Y Q.,2010. Wave Phenomena" and Formation of Excess Capacity. Econ. Res. J., 10,4-19.

Liu H, Shi J., 2013. Applying ARMA–GARCH approaches to forecasting short-term electricity prices. Energy Econ. 37, 152-166.

Liu M, Chen M, He G.,2017. The origin and prospect of billion-ton coal production capacity in China. Resor. Conserv. Reccy. 125,70-85.

Mathis S, Koscianski J., 1997. Excess capacity as a barrier to entry in the US titanium industry. Int. J. Ind. Organ. 15(2), 263-281.

Mulligan R F., 2016. The multifractal character of capacity utilization over the business cycle. Q. Rev. Econ. Finance. 63, 147-152.

Papadimitriou T, Gogas P, Stathakis E., 2014. Forecasting energy markets using support vector machines. Energy Econ. 44, 135-142.

Peel, M.C., Amirthanathan, G.E., Pegram, G.G.S., McMahon, T.A., Chiew, F.H.S., 2005. Issues with the application of empirical mode decomposition. In: Zerger, A., Argent, R.M. (Eds.), Modsim 2005 International Congress on Modelling and Simulation, pp. 1681–1687.

Pindyck R S., 1986. Irreversible investment, capacity choice, and the value of the firm. Am. Econ. Rev. 78(5), 969-985.

Pirard R, Irland L C., 2007. Missing links between timber scarcity and industrial overcapacity: Lessons from the Indonesian pulp and paper expansion. Forecast Policy Econ. 9(8), 1056-1070.

Ray S C., 2015. Nonparametric measures of scale economies and capacity utilization. Eur. J. Oper. Res. 245(2), 602-611.

Suganthi L, Samuel A A., 2012. Energy models for demand forecasting—A review. Renew. Sust. Enger. Rev. 16(2), 1223-1240.

Tang E, Peng C., 2017. A macro-and microeconomic analysis of coal production in China. Resour. Pol. 51, 234-242.

Tang, X., et al., 2016. China' s coal consumption declining—Impermanent or permanent? Resour. Conserv. Recy. http://dx.doi.org/10.1016/j.resconrec.2016.07.018

Tsai S B., 2016. Using grey models for forecasting China's growth trends in renewable energy consumption. Clean Technol. Envir. 18(2), 563-571.

Wang W C, Kwokwing C, Xu D M, et al., 2015. Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition. Water Resour. Manag. 29(8), 2655-2675.

Wang Y, Luo G, Guo Y., 2014. Why is there overcapacity in China's PV industry in its early growth stage? Renew. Energy. 72, 188-194.

Wang Y, Wang J, Zhao G, et al., 2012. Application of residual modification approach in seasonal ARIMA for electricity demand forecasting: A case study of China. Energy Policy 48, 284-294.

Wang, D., et al., 2016. Scenario prediction of China's coal production capacity based on system dynamics model. Resour. Conserv. Recy. http://dx.doi.org/10.1016/j.resconrec.2016.07.013.

Wang, D., Ma, G., Song, X., Liu, Y., 2017. Energy price slump and policy response in the coal-chemical industry district: A case study of Ordos with a system dynamics model. Energy Policy 104, 325-339.

Wu D, Yang H, Chen X, et al., 2008. Application of image texture for the sorting of tea categories using multi-spectral imaging technique and support vector machine. J. Food.Eng. 88(4), 474-483.

Wu K, Li B., 2015. Energy development in China national policies and regional strategies. Energy Policy 23(2), 167-178.

Wu X, Mao J, Du Z, et al., 2013. Online training algorithms based single multiplicative neuron model for energy consumption forecasting. Energy 59, 126-132.

Wu Z, Huang N E., 2009. Ensemble empirical mode decomposition: a noise-assisted data analysis method. Adv. Adapt. Data Anal. 1(01), 1-41.

Xie N, Yuan C, Yang Y., 2015. Forecasting China's energy demand and self-sufficiency rate by grey forecasting model and Markov model. Int. J. Electr. Power Energ. Syst. 66, 1-8.

Xiong T, Bao Y, Hu, Z., 2014. Does restraining end effect matter in EMD-based modeling framework for time series prediction? Some experimental evidences. Neurocomputing 123, 174-184.

Xu W, Gu R, Liu Y, et al., 2015. Forecasting energy consumption using a new GM–ARMA model based on HP filter: The case of Guangdong Province of China. Econ. Model 45, 127-135.

Yang Z, Ce L, Lian L., 2017. Electricity price forecasting by a hybrid model, combining wavelet transform, ARMA and kernel-based extreme learning machine methods. Appl. Energy 190, 291-305.

Yu L, Zhang J., 2014. Fundamental Cause and Solution of China's Excess Capacity: Non-market Factors and Three-Step-Strategy. Reform (2), 005.

Yu, S., Wei, Y.M., Wang, K., 2012. A PSO–GA optimal model to estimate primary energy demand of China. Energy Policy 42, 329-340.

Yuan J, Li P, Wang Y, et al., 2016. Coal power overcapacity and investment bubble in China during 2015–2020. Energy Policy 97, 136-144.

Zhang G, Wu Y, Liu Y., 2014. An advanced wind speed multi-step ahead forecasting approach with characteristic component analysis. J. Renew. Sustain. Ener. 6(5), 1663-1672.

Zhang H, Zheng Y, Ozturk U A, et al., 2016. The impact of subsidies on overcapacity: A comparison of wind and solar energy companies in China. Energy 94, 821-827.

Zhang J L, Zhang Y J, Zhang L., 2015. A novel hybrid method for crude oil price forecasting. Energy Econ. 49, 649-659.

Zhang W, Cheng Z, Zhou D, et al., 2014. Over-investment, diversification and local government intervention. Econ. Rev. 3, 139-152.

Zhang X, Lai K K, Wang S Y., 2008. A new approach for crude oil price analysis based on Empirical Mode Decomposition. Energy Econ. 30(3), 905-918.

Zhang Y F., 2014. Formation Mechanism and Regulation Measures Research on China Coal Overcapacity. Working Paper, China University of Mining and Technology.

Zhang Y, Nie R, Shi R, et al., 2016. Measuring the capacity utilization of the coal sector and its decoupling with economic growth in China's supply-side reform. Resour. Conserv. Recycl. http://dx.doi.org/10.1016/j.resconrec.2016.09.022

Zhang Y, Zhang M, Liu Y, et al., 2017. Enterprise investment, local government intervention and coal overcapacity: The case of China. Energy Policy 101, 162-169.

Zhao B F, Huang Z G., 2014. Reasearch on measurement and evolution characteristic of capacity utilization in coal industry of China. Stat. Info. Forum. 168(9), 15-20.

Zheng R., 2015. Analysis on causes and countermeasures of present overcapacity in China coal industry. Coal Econ. Res. 12, 48-52.

Zhu B, Wei Y., 2013. Carbon price forecasting with a novel hybrid ARIMA and least squares support vector machines methodology. Omega 41(3), 517-524.

Highlights

- We propose an integrated methodology to analyze and predict the coal overcapacity.
- The proposed method significantly outperforms other widely developed baselines.
- Institutional distortion is the most important cause of coal overcapacity in China.
- The impact of institutional distortion on coal overcapacity will be further exacerbated in future.
- From 2017 to 2019, the scale of coal overcapacity in China will be between 1.721 and 1.819 billion tons.