Energy Economics 70 (2018) 143–157

Contents lists available at ScienceDirect

Energy Economics

journal homepage: www.elsevier.com/locate/eneeco

A novel multiscale nonlinear ensemble leaning paradigm for carbon price forecasting

Bangzhu Zhu [a,*], Shunxin Ye [a], Ping Wang [a], Kaijian He [b,*], Tao Zhang [c], Yi-Ming Wei [d,*]

[a] School of Management, Jinan University, Guangzhou, Guangdong 510632, China
[b] School of Business, Hunan University of Science and Technology, Xiangtan, Hunan 411201, China
[c] Business School, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK
[d] Center for Energy and Environmental Policy Research, Beijing Institute of Technology, Beijing 100081, China

ARTICLE INFO

Article history:
Received 17 May 2017
Received in revised form 23 December 2017
Accepted 28 December 2017
Available online 06 January 2018

JEL Classification:
C45
C52
C53
E37
Q47
Q56

Keywords:
Carbon price forecasting
Least squares support vector machine
Empirical mode decomposition
Particle swarm optimization
Kernel function prototype

ABSTRACT

In this study, a novel multiscale nonlinear ensemble leaning paradigm incorporating empirical mode decomposition (EMD) and least square support vector machine (LSSVM) with kernel function prototype is proposed for carbon price forecasting. The EMD algorithm is used to decompose the carbon price into simple intrinsic mode functions (IMFs) and one residue, which are identified as the components of high frequency, low frequency and trend by using the Lempel-Ziv complexity algorithm. The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model is used to forecast the high frequency IMFs with ARCH effects. The LSSVM model with kernel function prototype is employed to forecast the high frequency IMFs without ARCH effects, the low frequency and trend components. The forecasting values of all the components are aggregated into the ones of original carbon price by the LSSVM with kernel function prototype-based nonlinear ensemble approach. Furthermore, particle swarm optimization is used for model selections of the LSSVM with kernel function prototype. Taking the popular prediction methods as benchmarks, the empirical analysis demonstrates that the proposed model can achieve higher level and directional predictions and higher robustness. The findings show that the proposed model seems an advanced approach for predicting the high nonstationary, nonlinear and irregular carbon price.

1. Introduction

With its increasingly severe consequences, global climate change has become a serious threat to human sustainable development. Carbon market, as an effective mechanism dealing with the global climate change, attracts significant attention of governments and organizations worldwide. In the past few years, global carbon market, represented by the European Union Emissions Trading System (EU ETS), has witnessed a rapid development. However, carbon price violent variations remarkably impact emissions reduction performances and market values. Predicting carbon price accurately can on the one hand enable us to understand the changeable patterns of carbon price and develop an efficient stabilization mechanism for carbon price, on the other hand help investors avoid carbon market risks and increase the value of carbon assets (Zhu et al., 2016). However, as emerging policy-based artificial markets, carbon markets are impacted by both internal market mechanisms and external environmental heterogeneity (Zhang and Wei, 2010), which causes the nonstationary and nonlinear characteristics of carbon prices. Therefore carbon price prediction presents a great challenge to researchers and becomes one of the priority topics in the fields of energy and climate economics.

In the literature there are various methods adopted for carbon price forecasting. Early studies mainly used the qualitative analysis to predict carbon price (Reilly and Paltsev, 2005; Kanen, 2006). Recent studies used more complex methods for carbon price forecasting, which can roughly be divided into three categories: statistical and econometric models, artificial intelligence models, and ensemble (hybrid) models. The traditional statistical and econometric models were widely used for carbon price forecasting. Chevallier (2011) used a nonparametric method to predict carbon prices, and found that the method could reduce the prediction error by almost 15% compared with linear autoregression models. Byun and Cho (2013) used the Generalized Autoregressive Conditional Heteroskedasticity (GARCH)-family models to predict carbon prices, and found that GJR-GARCH was more effective than TGARCH and GARCH in their particular case. Koop and Tole (2013) used the dynamic model averaging (DMA) method to forecast carbon prices, and obtained a high

* Corresponding authors.
E-mail addresses: wpzbz@126.com (B. Zhu), kaijian.he@my.cityu.edu.hk (K. He), ymwei@263.net (Y.-M. Wei).

https://doi.org/10.1016/j.eneco.2017.12.030
0140-9883/

prediction accuracy. María et al. (2015) applied the ARMAX-GARCH model with a time-varying jump probability to predict carbon prices, and obtained higher prediction accuracy than the standard ARMAX-GARCH model. Although these models based on stationary data and linear hypotheses can obtain high prediction accuracy, they cannot effectively deal with the nonlinearity in carbon price changes. To cope with this major limitation of statistical and econometric models, artificial intelligence models, including artificial neural networks (ANN), support vector machines (SVM) and least square SVM (LSSVM) were applied for carbon price prediction. For example, Zhu and Wei (2011) used the LSSVM method to predict carbon prices, and obtained better results than ARIMA and ANN. Fan et al. (2015) used the multi layered perceptron (MLP)-ANN model to predict carbon prices, and gained a good performance. These artificial intelligence models can effectively capture the nonlinear patterns hidden in carbon price variations, thus obtaining better prediction results than statistical and econometric models. In the meantime, in order to overcome the drawbacks of single models and further improve the accuracy of carbon price prediction, ensemble (hybrid) models were adopted for carbon price forecasting. Zhu (2012) used an empirical mode decomposition (EMD)-based ANN to predict carbon prices, and obtained a higher prediction accuracy than single ANN model. Zhu and Wei (2013) used a hybrid forecasting model, incorporating ARIMA and LSSVM, for carbon price forecasting, and found that the hybrid model was more effective than single ANN and ARIMA models. Wei and Can (2015) used the EMD–GARCH method to forecast the carbon prices of five pilots carbon markets in China, and gained the alternative interval of a lower bound of 30 yuan/ton and an upper bound of 50 yuan/ton in the national carbon trading market. Zhu et al. (2016) used an ensemble EMD (EEMD)-based LSSVM to predict the carbon price, and found that the proposed method could obtain high level and directional predictions. Atsalakis (2016) proposed a computational intelligence model with a novel hybrid neuro-fuzzy controller that forms a closed-loop feedback mechanism (PATSOS) to forecasting the daily carbon price, and get a higher accurate. Sun et al. (2016) combined variational mode decomposition (VMD) and spiking neural networks (SNNs) to improve forecasting accuracy and reliability. Zhu et al. (2017) used EMD and evolutionary LSSVM to improve the robustness of carbon price prediction, and obtained more robust performance than the other popular forecasting methods.

Extant studies have shown that multiscale ensemble forecasting can decompose the complex carbon price into several simple modes, so as to significantly improve the prediction accuracy of carbon price. However, there are several major drawbacks in existing studies. Firstly, in existent multiscale ensemble prediction, the same model is used to predict all the simple modes. As each mode has its own data characteristics, models with more appropriate mode-specific assumptions model is normally required for predicting it (Zhang et al., 2008; Zhu et al., 2015, 2016). Secondly, although LSSVM has a high nonlinear modeling ability, its predictive power is subject to model selection. Moreover, existing studies have predetermined the radical basis function (RBF) as the kernel function so as to search the model parameters of LSSVM (Yu et al., 2009; Zhu and Wei, 2013; Silva et al., 2015; Zhang et al., 2015). Few works have investigated the suitability of the kernel function for a specific problem. The inappropriate kernel function would negatively affect the accuracy of carbon price prediction. Thirdly, existing multiscale ensemble prediction models are limited to linear ensemble form, i.e. aggregating the forecasted values of all the modes into the forecasting ones of the original carbon price. A linear ensemble approach is not always appropriate for all the circumstances (Liao and Tsao, 2006; Alonso et al., 2007), thus affecting the accuracy of carbon price forecasting.

To address the existing drawbacks of carbon price prediction, this research develops a novel multiscale nonlinear ensemble leaning paradigm incorporating EMD, LSSVM with kernel function prototype, and particle swarm optimization (PSO) to improve the forecasting accuracy of nonstationary and nonlinear carbon price. It is a breakthrough study, making both methodological and empirical innovations. *Methodologically*, it develops a novel adaptive multiscale nonlinear ensemble learning paradigm for carbon price forecasting. Firstly, the EMD algorithm is used to decompose the carbon price into simple modes. Secondly, the obtained simple modes are identified as the components of high frequency, low frequency and trend by using the Lempel-Ziv complexity algorithm. GARCH can then be applied to predict the high frequency components with ARCH effects because of its strong short-term memory; and LSSVM with a universal kernel function prototype, is applied to forecast the high frequency components without ARCH effects, the low frequency and trend components. The proposed model can adaptively select the optimal kernel function type and model parameters according to the specific data using PSO, which can make good use of various kernel functions types and overcome the drawbacks of single kernel function. Finally, the LSSVM-based nonlinear ensemble approach is used to aggregate the prediction values of all the components acquired by different models into the forecasting values of the original carbon price, so as to further improve the prediction accuracy. *Empirically*, the proposed multiscale nonlinear ensemble learning paradigm has been tested with the data of the daily European Union Allowance futures prices from January 2, 2013 to April 14, 2015, obtained from the Intercontinental Exchange (ICE). Compared with popular prediction methods, the empirical analysis shows that the proposed model is the optimal, and can effectively deal with the nonlinearity and nonstationarity of carbon prices.

The rest of the paper is organized as follows: Sections 2.1, 2.2, 2.3 and 2.4 construct the novel adaptive multiscale nonlinear ensemble learning paradigm. Section 3 conducts empirical analysis. Section 4 concludes the study.

2. Methodology

2.1. Empirical mode decomposition

EMD was proposed as a multiscale decomposition technique that takes advantage of the local characteristics scales of the underlying data components and extracts these components known as Intrinsic Modes from the data. It was originally proposed by Huang et al. (1998) as an effective empirical method for the nonlinear and non-stationary data. These intrinsic modes are defined as the intrinsic mode function (IMF), satisfying the following conditions: (i) the difference between the extrema and zero-crossings do not exceed 1. (ii) The functions are zero mean locally and symmetric. IMFs satisfying these conditions are zero mean and nearly periodic. Thus they are harmonic with changing amplitudes and frequencies at different timescales (Huang et al., 1999).

Compared with the traditional Fourier and wavelet decompositions, EMD technique has several distinct advantages. Firstly, it is relatively easy to understand and implement EMD. Secondly, since the decomposition is based on the local characteristic timescales of the data and only extrema are used in the sifting process, EMD is local, self-adaptive, and very implicative. It is highly efficient for nonlinear and nonstationary time series decomposition, therefore can adaptively and robustly decompose carbon price time series into several IMFs and one residue that display linear and nonlinear behaviors depend only on the nature of carbon prices. Thirdly, the IMFs derived from EMD have a clear instantaneous frequency as the derivative of the phase function. Thus the Hilbert transformation can be applied to the IMFs, allowing us to explore the data in a time-frequency-energy space. Last but not the least, in wavelet decomposition, a filter base function must be determined beforehand. However, it is difficult for some unknown series to determine the filter base function. Unlike wavelet decomposition, EMD does not have to determine a filter base function before decomposition. The four merits discussed above makes EMD an effective decomposition tool.

2.2. LSSVM with kernel function prototype

A kernel function is the critical parameter for the LSSVM predictor. When applied in practice, the kernel function used in LSSVM predictor