Energy Economics 71 (2018) 201–212

Contents lists available at ScienceDirect

Energy Economics

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

A novel approach for oil price forecasting based on data fluctuation network

Minggang Wang [a,b], Lixin Tian [a,*], Peng Zhou [c]

[a] School of Mathematical Science, Nanjing Normal University, Nanjing 210042, Jiangsu, China
[b] Department of Mathematics, Taizhou College, Nanjing Normal University, Taizhou 225300, Jiangsu, China
[c] College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, Jiangsu, China

ARTICLE INFO

Article history:
Received 19 June 2017
Received in revised form 14 February 2018
Accepted 23 February 2018
Available online 01 March 2018

JEL classifications:
C45
C53
Q47
C22
C02
C15

Keywords:
Time series
Complex network
Oil price
Prediction

ABSTRACT

Characterizing nonlinear time series using complex network science is a new multidisciplinary methodology. This paper puts forward a new time series prediction method based on data fluctuation network, named data fluctuation networks predictive model (DFNPM). The basic idea of the method is: first map time series into data fluctuation network and extract the fluctuation features of time series according to the topological structure of the networks, and then construct models with useful information extracted to predict time series. With Cushing, OK Crude Oil Future Contract 1 (Dollars per Barrel) and New York Harbor Regular Gasoline Future Contract 1 (Dollars per Gallon) as its sample data as well as DFNPM as its prediction model, the research makes a prediction on crude oil and gasoline futures prices from December 30, 2014 to February 26, 2015. A comparison is conducted between the result of the prediction and such traditional prediction models as grey prediction (GM) model, exponential smoothing model (ESM), autoregressive integrated moving average (ARIMA) model and radial basis function neural network (RBF) model, which shows that DFNPM performs significantly better than the above four traditional prediction models in both the direction and level of prediction.

1. Introduction

The fluctuation of oil price plays an important role in social economy. Therefore, the prediction of oil price has always been a hot research topic for scholars at home and abroad. Some oil price prediction methods are based on traditional econometric models. For instance, Xiang and Zhuang (2013) used an autoregressive integrated moving average (ARIMA) model to predic the Brent monthly crude oil price from November 2012 to April 2013. Nomikos and Andriosopoulos (2012) estimated the conditional mean and volatility of West TexasIntermediate (WTI) daily crude oil spot prices from December 9, 2000 to January 2, 2010, based on the generalized autoregressive conditional heteroskedasticity (GARCH) family models. Murat and Tokat (2009) used the random walk (RW) model to forecast oil price movements and crack spread futures. Mirmirani and Li (2005) used the vector autoregression (VAR) model to predict the US monthly oil price covering the period from January 1980 to November 2002. Lanza et al. (2005) used the error correction model (ECM) to predict the WTI and Brent weekly crude oil prices with the sample period from 1994 to 2002. Morana (2001) proposed a semiparametric approach to oil price forecasting, which allows one to forecast the entire oil price distribution at different time horizons, without requiring the specification of a structural model for the conditional mean of the oil price process. Other methods are based on artificial intelligence (AI) and can effectively distinguish random factors which traditional econometric models are unable to do. These AI methods can provide more accurate results for some situations. For example, Movagharnejad et al. (2011) introduced the artificial neural network (ANN) to forecast the quantitative data of crude oil prices over the period from January 2000 to April 2010. Chiroma et al. (2015) presented an evolutionary neural network (ENN) to predict the WTI monthly crude oil price data from May 1987 to December 2011. Khashman and Nwulu (2011) predicted the WTI weekly spot crude oil prices from January 03, 1986 to December 25, 2009, based on the SVR model. Similarly, Li and Ge (2013) predicted the crude oil prices from May 1994 to December 1995 based on an e-SVR model with dynamic errors correction. Xie et al. (2006) compared the support vector regression (SVR) with the ARIMA and back-propagation neural network (BPNN), and witnessed the superiority of the SVR in the prediction for the WTI monthly prices from January 1970 to December 2003. Li et al. (2013) predicted the WTI weekly data from January 4, 2008 to October 18, 2013 using least squares support vector machines

* Corresponding author.
E-mail addresses: tianlx@ujs.edu.cn (L. Tian), cemzp@nuaa.edu.cn (P. Zhou).

https://doi.org/10.1016/j.eneco.2018.02.021
0140-9883/

(LSSVM), and argued that the LSSVR outperformed the ARIMA, SVR and BPNN models. However, these AI models have their own weaknesses, e.g., parameter sensitiveness and potential over-fitting (Tang et al., 2012). Recently, the hybrid prediction methods have developed and become a predominant type for oil price forecasting. For instance, Wang et al. (2005) proposed the promising concept of decomposition and ensemble, and then a series of decomposition–ensemble learning paradigms have been developed for crude oil price forecasting. In a typical decomposition–ensemble model, three main steps are included, i.e., data decomposition for simplifying the complex data, individual prediction for each decomposed mode, and ensemble prediction for final prediction result (Tang et al., 2012; Yu et al., 2008; Liu et al., 2013; Zhang et al., 2008). Zhang et al. (2015) propose a hybrid method that combines ensemble empirical mode decomposition (EEMD), least square support vector machine particle swarm optimization (LSSVM-PSO), and the GARCH model to forecast crude oil prices, and find this hybrid method has a strong crude oil price forecasting capability.

As mentioned above, the researchers have developed many prediction models to carry out prediction on oil price. The past literatures have shown that oil price forecasting results are sensitive to the modeling sample interval selection, sample data frequency and sample structural breaks etc. (Xiang and Zhuang, 2013; Nomikos and Andriosopoulos, 2012; Murat and Tokat, 2009; Mirmirani and Li, 2005; Lanza et al., 2005; Morana, 2001; Movagharnejad et al., 2011; Chiroma et al., 2015; Khashman and Nwulu, 2011; Li and Ge, 2013; Xie et al., 2006; Li et al., 2013; Tang et al., 2012; Wang et al., 2005; Yu et al., 2008; Liu et al., 2013; Zhang et al., 2008, 2015). Up to now, oil price prediction remains an open question, for the main reason that on one hand, the fluctuation of oil price time series is affected by various factors in market, such as political environment, natural environment and personal psychology. The interaction and collision among different factors bestow oil price time series with features of strong noise, instability and invisible periodicity. On the other hand, there are more or less defects in the existing prediction methods. For example, prediction model construction from the perspective of traditional statistics should be based on the stability in a statistical sense, or it may produce relatively poor prediction results for nonlinear time series (Xiang and Zhuang, 2013; Nomikos and Andriosopoulos, 2012; Murat and Tokat, 2009; Lanza et al., 2005; Morana, 2001). Although artificial intelligence (AI) method can well handle nonlinear prediction, there are some downsides for the method like low convergence rate, difficulty in choosing parameters and high likelihood to fall into local minimum (Movagharnejad et al., 2011; Chiroma et al., 2015; Khashman and Nwulu, 2011; Li et al., 2013; Xie et al., 2006; Tang et al., 2012). Therefore, there are still many problems worth discussion in the field of oil price prediction.

The prediction of oil price is essentially a kind of time series forecasting problem. Time series prediction has been a classic issue in the field of time series research. At present, there are mainly three types of time series prediction methods. First, statistical method based on traditional measurement theories. From the perspective of statistics, time series prediction and analysis, in nature, is a process in which a series of data is arranged according to the fluctuation over time and random factors of time series are extracted by statistical method to find out the interdependent relationship among data. The relationship is expressed in mathematical language to predict future data fluctuation. There are still many economic time series prediction methods within the range of statistics, including random walk (RW) theory (Kac, 1947), moving average (MA) theory (Box and Pierce, 1970), autoregressive (AR) model (Dijk et al., 2002), autoregressive moving average (ARMA) model, autoregressive integrated moving average (ARIMA) model, autoregressive conditional heteroskedasticity (ARCH) model and generalized autoregressive conditional heteroskedasticity (GARCH) model (De Gooijer and Kumar, 1992; Makridakis, 1993). Second, intelligent computing method based on computer science. The intelligent computing method can effectively determine random and complicated factors which traditional statistical method is unable to find out. Generally speaking, prediction by intelligent computing method promises a more accurate result. The classic intelligent computing techniques include artificial neural network (ANN) (Moody and Darken, 1989), support vector machine (SVM) (Boser et al., 1992; Mika et al., 1999) and the evolutionary algorithm (EA) (Tay and Cao, 2002; Niu et al., 2010). Third, the hybrid prediction methods which combine the advantages of various kinds of single prediction methods. In terms of complicated and nonlinear data, a single prediction method can only take into consideration some major factors while has no way to cover comprehensive and effective information of time series change, thus leading to low accuracy of the prediction results. Therefore, the hybrid prediction method (Winkler, 1989; Wang, 2009; Wang et al., 2012) has attracted more and more attention because of its advantages in solving the problem. Actually, the hybrid prediction method is a process in which the hidden information of series to be predicted is subdivided into a series of components and different prediction methods are chosen for information extraction according to the features of the components. In this way, the uncertainty caused by a single method can be reduced, so as to improve the control level of the whole prediction process.

In fact, to ensure high accuracy of oil price prediction, it is a must to first draw out precisely the regular pattern of the fluctuation of oil price time series and then extracts effectively the useful information of the fluctuation. In recent years, extensive attention has been cast to the application of complex network theory to nonlinear time series analysis. Its major concept is to map the nonlinear time series into corresponding complex networks through certain algorithms and then draw out the regular pattern for the fluctuation of nonlinear time series through the typology of complex networks. A lot of studies show that the application of complex network theory can effectively help find out the characteristics of time series, thus giving rise to many new algorithms that map time series into complex networks, including visibility graph (VG) (Lacasa et al., 2008), pseudo-periodic time series transform algorithm (Xu et al., 2008), phase space reconstruction method (Zhang and Small, 2006) and coarse graining method of phase space (Wang et al., 2017). Currently, the researchers have introduced the time series complex network analysis technique to the field of energy economy to study the fluctuation features of energy price (Wang and Tian, 2016; Chen et al., 2010; An et al., 2014; Wang et al., 2016b, 2016a). For example, Chen et al. (2010) constructs the network of international oil price to find out the dynamics features of the network. An et al. (2014) discusses the autoregressive effects of the fluctuation of crude oil price time series. Wang et al. (2016a) construct a directed and weighted network for international crude oil and gasoline prices to analyze the fluctuation features of international crude oil and gasoline prices in different periods. The above mentioned studies, through making use of complex network theory to study the fluctuation features of energy price, yield many valuable results.

Based on the previous studies, this paper uses time series complex network analysis techniques to put forward a new prediction method for oil price. This method is based on that converts oil price fluctuation sequences into complex networks, so this method is called data fluctuation networks predictive model (DFNPM). The basic approach is: transform oil price fluctuation time series into a directed and weighted network through certain algorithms and extract fluctuation features of oil price according to the topological structure of the network, and then construct models with useful information extracted to predict time series. We organize the rest of this paper as follows. Section 2 provides a detailed description of how the proposed model was formulated. Section 3 describes and discusses the oil forecasting results. At the end of the paper we present our conclusions and propose possible future lines of research.

2. Model construction

By means of coarse graining method (Wang et al., 2016a) mapped oil price fluctuation series into a character string composed of five