International Journal of Forecasting 34 (2018) 665–677

Contents lists available at ScienceDirect

International Journal of Forecasting

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

Crude oil price forecasting based on internet concern using an extreme learning machine

Jue Wang [a,b,*], George Athanasopoulos [c], Rob J. Hyndman [c], Shouyang Wang [a,b]

[a] School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China
[b] CEFS, MADIS, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
[c] Department of Econometrics and Business Statistics, Monash University, Australia

ARTICLE INFO

Keywords:
Crude oil futures price
Internet concern
BEMD
ELM

ABSTRACT

The growing internet concern (IC) over the crude oil market and related events influences market trading, thus creating further instability within the oil market itself. We propose a modeling framework for analyzing the effects of IC on the oil market and for predicting the price volatility of crude oil's futures market. This novel approach decomposes the original time series into intrinsic modes at different time scales using bivariate empirical mode decomposition (BEMD). The relationship between the oil price volatility and IC at an individual frequency is investigated. By utilizing decomposed intrinsic modes as specified characteristics, we also construct extreme learning machine (ELM) models with variant forecasting schemes. The experimental results illustrate that ELM models that incorporate intrinsic modes and IC outperform the baseline ELM and other benchmarks at distinct horizons. Having the power to improve the accuracy of baseline models, internet searching is a practical way of quantifying investor attention, which can help to predict short-run price fluctuations in the oil market.

1. Introduction

International crude oil price prediction has become an increasingly important issue. Crude oil plays a significant role in the global economy (Uri, 1996), with the crude oil market accounting for nearly two-thirds of the world's energy demand (Alvarez-Ramirez, Soriano, Cisneros, & Suarez, 2003). A leap in the price of crude oil would result in inflationary pressure and an economic recession within most countries, and therefore would have a significant impact on the global economy. In contrast, a rapid decline in crude oil prices would prohibit economic development in oil-producing countries, thus creating political instability and potentially social unrest. As a result, it is necessary to investigate the inherent mechanisms of oil price fluctuations in order to reduce the potential risks of oil price volatility.

In general, crude oil prices are determined by supply and demand (Hagen, 1994; Stevens, 1995), although they are also influenced by speculation and extreme events, which can intensify the price volatility and market instability. Numerous studies (Demirer & Kutan, 2010; Kaiser & Yu, 2010; Zhang, Yu, Wang, & Lai, 2009) have argued that additional market factors should be intercalated into an analytical framework for the prediction of crude oil prices. Oil prices are sensitive to oil-related events such as war, extreme weather, OPEC production stipulations, etc. Ji and Guo (2015) investigated the effects of four types of oil-related events on world oil prices and concluded that oil prices respond differently to specific events. Hamilton (2009) highlighted the fact that the Iraq War and the Iranian Revolution resulted in oil supply disruptions that of course impacted oil prices.

* Corresponding author at: CEFS, MADIS, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China.
E-mail address: wjue@amss.ac.cn (J. Wang).

https://doi.org/10.1016/j.ijforecast.2018.03.009
0169-2070/

The recently introduced concept of internet concern (IC) involves the use of crude oil market search data for quantifying investor speculation. IC is now an important factor in explorations of the impact and magnitude of market concerns. Numerous studies have suggested that information extracted from the internet can contribute to the prediction of financial data (Bollen, Mao, & Zeng, 2011; Bordino et al., 2012). Health economists have used Google search queries to provide early indicators in uncovering disease incidence and prevalence rates, while others have used such queries to predict consumer spending. For example, Choi and Varian (2012) highlighted the fact that queries data can be useful for indicating consumers' planned purchasing. Other studies (e.g. Askitas & Zimmermann, 2009; Bank, Larch, & Peter, 2011; Preis, Moat, & Stanley, 2013) have used Google search data to measure IC in financial markets. Li, Ma, Wang, and Zhang (2015b) used a Google search volume index to quantify investor attention and investigated the relationships between search data, differential trader positions and crude oil prices. Park, Lee, and Song (2016) also utilized internet search data from Google Trends for forecasting the short-term flow of Japanese tourists to South Korea. Yao and Zhang (2017) explored the effects and predictive power of the Google Index on crude oil prices by incorporating the Google Index into ARIMA and ARMA-GARCH models as an exogenous variable.

Various traditional statistical and econometric models, such as cointegration, GARCH, vector autoregression (VAR) and Markov models, have been adopted for analyzing oil markets (Allegret, Mignon, & Sallenave, 2015; Maghyereh, 2006; Salisu & Oloko, 2015; Zhang & Wang, 2015). Considering the nonlinear patterns and irregularities hidden within the oil price series, artificial intelligence models such as neural networks (NN; see Yu, Wang, & Lai, 2008), support vector machines (SVM; see Xie, Yu, Xu, & Wang, 2006), and genetic algorithms (GA; see Motlaghi, Jalali, & Ahmadabadi, 2008), have also been used for forecasting crude oil prices. For instance, Chiroma, Abdulkareem, and Herawan (2015) proposed a hybrid approach based on a genetic algorithm and neural network (GA-NN) for predicting the West Texas Intermediate (WTI) crude oil price. Baruník and Malinska (2016) proposed the use of a generalized regression framework based on neural networks for forecasting oil prices. Single layer feed-forward networks (SLFN) are used widely in classification and regression analysis applications. However, SLFN and other such gradient descent learning methods are time-consuming and suffer significant critical errors, such as over-fitting, local minima, etc. Huang, Zhu, and Siew (2004, 2006) proposed a learning algorithm known as 'extreme learning machine' (ELM), which has performed well for predicting non-linear time series and has a better generalization performance than the gradient-based learning methods. ELM has been implemented widely for short-term wind power forecasting (Abdoos, 2016), hyperspectral imagery classification (Li, Chen, Su, & Du, 2015a), electricity price forecasting (Xiao et al., 2016), and online sequential prediction (Wang & Han, 2015).

In parallel to the development of forecasting technologies, numerous decomposition and construction methods have been developed, such as wavelet analysis (Shahbaz, Tiwari, & Tahir, 2015), singular spectral analysis (SSA; see Fenghua, Jihong, Zhifang, & Xu, 2014), and empirical mode decomposition (EMD; see Huang et al., 1998). EMD has been used as an effective analysis model in economics and finance (He, Zha, Wu, & Lai, 2016; Yu et al., 2008; Zhang et al., 2009). As an extension of EMD into the two-dimensional space, bivariate empirical mode decomposition (BEMD) was first proposed by Rilling, Flandrin, Gonalves, and Lilly (2007). BEMD simultaneously models the joint oscillating modes at each intrinsic mode function (IMF) and provides a robust estimate of asymmetry for nonlinear and nonstationary data (Molla, Ghosh, & Hirose, 2011; Yang, Court, Tavner, & Crabtree, 2011).

This paper analyzes internet attention on the crude oil market, along with the impacts of two oil-related events, namely abnormal climate incidents and war. Three IC indices for capturing the influences of internet attention are constructed and an ELM-based forecasting model is established that incorporates intrinsic modes and IC. The objectives are to:

1. quantify the influence of emergencies on the crude oil market using the information extracted from the internet;
2. examine the advantages of the BEMD-based modeling framework for analyzing the transmission between each IC index and oil price volatility, while characterizing the magnitude and dynamics of the impacts at various frequencies; and
3. investigate the power of forecasting models with the aid of intrinsic modes and internet data for short-run crude oil price volatility.

Section 2 presents the methodology formulation of the basic BEMD and ELM, while Section 3 presents the main contribution, namely an IC analysis of the crude oil market based on BEMD and price forecasting using the ELM model. Section 4 reports experimental results, and Section 5 provides some conclusions.

2. Preliminaries

2.1. Bivariate empirical mode decomposition

EMD is a signal processing technique that decomposes a univariate (real-valued) signal into waveforms by extracting all of the oscillatory modes embedded within the signal. The waveforms extracted by EMD are named intrinsic mode functions (IMF); these are modulated in both amplitude and frequency.

BEMD is a generalized extension of the EMD for complex signals, and is particularly suitable for estimating amplitude information simultaneously across different frequencies for two nonlinear and nonstationary time series. In BEMD, two variables are decomposed simultaneously based on their rotating properties.

Assumptions

We decompose the two-dimensional data using the BEMD technique by introducing the following assumptions: