Energy 46 (2012) 564–574

Contents lists available at SciVerse ScienceDirect

Energy

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

Crude oil price analysis and forecasting using wavelet decomposed ensemble model

Kaijian He [a,b,*], Lean Yu [a], Kin Keung Lai [b]

[a] School of Economics and Management, Beijing University of Chemical Technology, Beijing 100029, China
[b] Department of Management Sciences, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong

ARTICLE INFO

Article history:
Received 22 November 2011
Received in revised form
26 July 2012
Accepted 28 July 2012
Available online 30 August 2012

Jels classification:
C22
C53
Q47

Keywords:
Wavelet analysis
Crude oil price forecasting
Ensemble algorithm
ARMA model
Multiresolution analysis

ABSTRACT

To improve the forecasting accuracy of crude oil price with deeper understanding of the market microstructure, this paper proposes a wavelet decomposed ensemble model. The proposed model follows the Heterogeneous Market Hypothesis that assumes the unstationarity and dynamic changing nature of the underlying market structure and introduces the wavelet analysis to analyze the dynamic underlying Data Generating Process at finer time scale domain. The simple averaging based ensemble model is introduced to reduce the estimation bias resulting from the use of different wavelet families by deriving market consensus view. The ensemble members are selected dynamically based on their in-sample performance among forecast matrices based on different wavelet families. Results from empirical studies show the superior performance of the proposed algorithm against the benchmark models, in terms of both level and directional predictive accuracy. The proposed model can effectively extract and model the time varying heterogeneous market microstructure, whose accurate characterization results in further improvement in market analysis and predictability.

1. Introduction

As the backbone of the world economy, crude oil has penetrating effects on different parts of the society. The forecasting of its price movement has been of critical concern to both industries and academics worldwide. From industry perspective, the forecasting of its movement is an integrated part of the decision making process, from the valuation, exploration to the development and production process. It influences different operational decision making process and the hedging behaviors of individual firms in wide range of industries such as exporting, manufacturing, and transportation. From government perspective, its evaluation and forecasting affect various short term and long term decision making process for export policy and national reserves. For academics community, it is closely related to some important theoretical issues in the economics and finance fields such as Efficient Market Hypothesis (EMH) and the Derivative Pricing [1,2]. Due to the fundamental and prominent roles the crude oil holds, the research has attracted continuous research attention over years.

In recent years, with the accelerating structural changes such as deregulation wave in the crude oil market and the globalization trend with the help of technological development such as widespread use of electronic trading platform, we have witnessed more diversified participants in the market, with increasingly versatile trading behaviors. This, along with the traditional risk sources including the relative rate of growth (changes) of supply and demand, competitions (competitors' pricing decisions), commodities substitutions, technology trends and international market conditions, as well as uncertainties arising from political instabilities and wars and conflicts in the oil-rich areas, has contributed to a more competitive and volatile environment in the crude oil market. Facing these changing market conditions, risk management needs from participants stimulate innovations in financial engineering and risk measurement field such as derivative, hedging and Value at Risk, whose efficient and effective usage depends on the deeper understanding and more accurate forecasting of crude oil price movement. Thus forecasting remains one of the most challenging issues in the crude oil field and has been under

* Corresponding author. School of Economics and Management, Beijing University of Chemical Technology, Beijing 100029, China. Tel./fax: +86 (0)10 64438793.
E-mail addresses: paulhekj@gmail.com (K. He), yulean@amss.ac.cn (L. Yu), mskklai@cityu.edu.hk (K.K. Lai).

0360-5442/$ – see front matter
http://dx.doi.org/10.1016/j.energy.2012.07.055

extensive investigations over years. Several lines of researches have been carried out to incorporate nonlinearity, autocorrelation and heteroscedasticity data features into the modeling process, aiming at improving the forecasting accuracy further. Results to date are largely inconclusive in the literature. Linear econometric model generally fails to provide satisfactory prediction accuracy, beyond the traditional Random Walk (RW) and Autoregressive Moving Average (ARMA) models. For example, the equilibrium model analyzes the economic relationship among participants in the physical crude oil market and derives the analytic equations to model them. Typical models including the supply and demand equilibrium model and more recently the co-integration model. For example, Abosedra and Baghestani [3] tested the effectiveness of futures price forecasting and found it to be superior to the Random Walk (RW) model [3]. Ye et al. [4] and Ye et al. [5] incorporated inventory variables into single equation models, firstly as a single variable and then as both high and low inventory variables, and improved model fit and out-of-sample forecasting accuracy [4,5]. Recently Ye et al. [6] incorporated further the cumulative excess production capacity as the ratcher variable into forecasting models and showed that it results in improved performance [6]. Chevillon and Rifflart [7] used two co-integration OPEC related relationships in the market and developed a forecasting equation based on it [7]. Interestingly Alquist and Kilian [8] show contradictory results. They found RW model to be by far the best models available [8]. Cuaresma et al. [9] derive a simple unobserved component model incorporating asymmetric cycles and found it to be superior in performance than the symmetric counterparts and benchmark Autoregressive (AR) models [9]. Knetsch [10] proposed a forecasting model using the present value model of rational commodity pricing and found it to have superior performance than the RW model [10]. Due to the inconclusive results in the literature, both RW and ARMA models serve as important benchmark in crude oil price forecasting and this practice has been witnessed in numerous works in the literature such as Alquist and Kilian [8], He et al. [11], Ye et al. [4], Yu et al. [12], Abosedra and Baghestani [3], Knetsch [10], Moshiri and Foroutan [13], Shambora and Rossiter [14].

Meanwhile, empirical work utilizing the power of an artificial intelligence model is on the rise, but with mixed results. Yu et al. [12] found that Artificial Neural Network (ANN) outperforms ARMA model, but has room for further improvement using ensemble algorithms. Fan et al. [15] proposed the Generalized Pattern Matching model based on genetic algorithm (GPMGA), which resulted in improved performance compared with a Recurrent Neural Network (RNN) based approach and Pattern Modeling and Recognition System (PMRS) [15]. Shambora and Rossiter [14] showed that ANN provides superior performance in terms of extensive performance measures including cumulative returns, year-to-year returns, returns over a market cycle and Sharpe ratios, than benchmark models such as RW model. Moshiri and Foroutan [13] proposed a neural network model to model the unknown nonlinear economic structure and demonstrate its effectiveness in performance improvement in empirical studies using crude oil futures prices [13]. Although numerous researches report positive results for the performance improvement from employing neural network to model the nonlinearity data patterns, the performance improvement is not consistent for all test cases [16]. Another issue raised with the employed computational based approaches is their ‘black box’ approach with the instability issue due to its inherent overfitting and local maximum issues [12,17,18].

In recent years, empirical researches suggest that the crude oil prices exhibit multiscale behavior, that is consistent with the Heterogeneous Market Hypothesis (HMH). Market is heterogeneous in nature, so that exploitation of this stylized fact leads to better understanding of the underlying Data Generating Processes (DGPs) and forecasting accuracy. Therefore, we move away from traditional parametric and artificial intelligence techniques based approach to introduce a novel semi-parametric multiscale wavelet based ensemble paradigm to analyze and model the dynamic DGPs at finer time scale domain, providing insights into different aspects of the market microstructure. Firstly, the wavelet analysis is introduced to model the multiscale market structure. Based on different wavelet families, the data are decomposed at different scales into sub data series that are dominated by simpler underlying DGPs to analyze and model the multiscale heterogeneous market structure. Then we further introduce the ensemble framework since individual forecasts are biased as influenced by different wavelet families employed and ensemble forecasts can help reduce the estimation biases. The proposed unifying paradigm incorporates different data characteristics such as autocorrelation, multiscale heterogeneity and parameter instability during the modeling process. Experiments based on crude oil market price series are conducted to evaluate and compare the performance of the proposed Wavelet Decomposed Ensemble (WDE) model against the benchmark RW and ARMA models using standard level and directional test statistics. Experiment results confirm that WDE model improves the level and directional forecasting accuracy. It also offers valuable information as to the underlying micro market structure.

Theoretical motivations for the current paper is the challenges to the fundamental Efficient Market Hypothesis (EMH) and the long standing question whether the market is predictable. The traditional EMH, even in its weak form, dictates that the market is not predictable. But this is stated in the equilibrium conditions in the long run, which doesn’t necessarily hold in the transitional periods when the market stabilizes from shocks in the short to medium time horizon. To help make the market operations more efficient, price predictability exists to a certain degree in the short to medium time horizon. But the micro-behavior of the market price has been largely ignored until fairly recently when the work on the Fractal Market Hypothesis (FMH) and HMH provided some further contending theories and evidence. The main research problem of this paper is mainly set within the general framework of HMH.

Practical motivation for the current paper is the challenge to improve forecasting and estimation accuracy as modeling accuracy has been difficult to improve beyond the simple RW levels. Recent empirical researches suggest that this is due to its complicated and nonlinear data nature, compared to the linear assumptions adopted before. Besides, crude oil markets typically represent a more volatile and risky environment than other commodities markets and the traditional financial markets [2,19,20]. It demonstrates unique characteristics due to several reasons. Firstly, crude oil is traded less frequently than equities. There are considerable transaction and opportunity costs involved, resulting in higher levels of fluctuations and relatively lower levels of efficiency. Secondly, price movements in crude oil markets are frequently influenced by numerous factors, such as weather conditions, political stability, economic prospects, consumer expectations and business indicators. The complex dynamics involved exhibit multiscale nonlinear characteristics, which are beyond explanatory capabilities of current models and methodologies for forecasting purposes. Despite its significance, forecasting of crude oil price has not received sufficient attention, as compared to other asset prices. Research in this paper attempts to address the unique characteristics of crude oil markets when constructing useful models for forecasting purposes and to fill the gap in the relevant literature.

This paper makes contributions in the following four aspects. Firstly, the HMH used in this paper receives less attention in the