Information Sciences 460–461 (2018) 279–292

Contents lists available at ScienceDirect

Information Sciences

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

A semi-heterogeneous approach to combining crude oil price forecasts

Jue Wang[a,b,*], Xiang Li[a,b], Tao Hong[c], Shouyang Wang[a,b]

[a] School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China
[b] CFS, MDIS, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
[c] Systems Engineering and Engineering Management Department, University of North Carolina at Charlotte, USA

ARTICLE INFO

Article history:
Received 13 January 2018
Revised 8 March 2018
Accepted 9 May 2018

Keywords:
Crude oil price
Decomposition and reconstruction
Forecast combination
Model selection

ABSTRACT

Crude oil price forecasting has received increased attentions due to its significant role in the global economy. Accurate crude oil price forecasts often lead to a rapid new production development with higher quality and less cost. Making such accurate forecasts, however, is challenging due to the intrinsic complexity of oil market mechanism. Many techniques have been tested in the crude oil price forecasting literature. Although forecast combination is a well-known method to improve forecast accuracy, generating forecasts using various techniques tend to be labor intensive. How to efficiently generate many individual forecasts for combination becomes a research question in crude oil price forecasting. Recently, several signal decomposition methods have been suggested for processing the oil price signals. In this paper, we propose a semi-heterogeneous approach to combining crude oil price forecasts, which interacts a set of decomposition methods with a set of forecasting techniques. We first decompose the original price series using four decomposition methods, such as Wavelet Analysis, Singular Spectral Analysis, Empirical Mode Decomposition, and Variational Mode Decomposition. We then use four different forecasting techniques, such as Autoregressive Models, Autoregressive Integrated Moving Average Models, Artificial Neural Networks, and Support Vector Regression Models, to forecast the components from each decomposition methods. Finally, we reconstruct the price forecasts from the forecasted components. This process generates 16 price forecasts in total for combination. We test the combination based on all individual forecasts, as well as a subset of the individual forecasts selected using Tabu Search. The experimental results demonstrate that the forecasting models with the addition of a decomposition technique can have an error reduction of 30.6% compared to benchmark models on average. The combined forecasts outperform the individual forecasts on average. Furthermore, comparing with the heterogeneous combination of 4 individual forecasts, the semi-heterogeneous combinations reduce the errors by 56.6% (w/o Tabu Search) and 61.6% (w/ Tabu Search).

* Corresponding author at: School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China.
E-mail address: wjue@amss.ac.cn (J. Wang).

https://doi.org/10.1016/j.ins.2018.05.026

1. Introduction

Crude oil is an important sector in the global economy. Crude oil prices are becoming increasingly volatile due to the stochastic nature of the global economy and financial policy. Accurate forecasts of primary commodity prices can help support the development of stabilization policies and budgetary planning in both oil-producing and oil-consuming countries [11].

The research community has tried many techniques for crude oil price forecasting. These techniques can be roughly categorized into two groups, econometric techniques and artificial intelligence. Among the econometric models that have been applied to crude oil price forecasting, such as Autoregressive models (AR), Autoregressive Integrated Moving Average (ARIMA) models, Vector Auto-regression (VAR) models and Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) models [3,24,29,37,48], AR and ARIMA are quite popular due to their ease of implementation and high accuracy. Many artificial intelligence techniques have been introduced to capture the nonlinear patterns and irregularities hidden in the oil price series. The popular ones include Support Vector Regression (SVR) and Artificial Neural Networks (ANN). For instance, Yu et al. introduced SVR and LSSVR learning paradigm to predict crude oil price [42,44]. Barunik et al. used ANN to forecast the term structure of crude oil futures prices [6].

Another strategy to forecast crude oil prices is to decompose the price series before building models. Many decomposition techniques can be adopted here, such as Wavelet Analysis (WT) [30], Empirical Mode Decomposition (EMD) [16,36], Singular Spectral Analysis (SSA) [23,38], and Variational Mode Decomposition (VMD) [12]. Reboredo et al. [28,30] used wavelet multiple decomposition to analyze the relationship between exchange rates and crude oil prices. Wavelet multiple decomposition has also been used to analyze the relationship between a range of macroeconomic variables and crude oil prices [1]. SSA was introduced to climatological statistical diagnosis and forecasting studies by Vautard et al. [35] in the late 1980's. EMD has been widely used to recursively decompose a signal into distinct modes separating spectral bands. EMD has also been shown to be an effective analysis model in economics and finance [14,21,45,46]. Unlike EMD, which is sensitive to noise and sampling methods, VMD appears to be robust to noise and sampling [12].

Many studies have argued that decomposition and reconstruction may help improve forecast accuracy [25,43]. Much effort to improve forecasting models was on the integrated technique of decomposition and reconstruction, such as integrating AR to EMD [13], as well as ARIMA to WT [19] and EMD [22]. Support vector machine (SVM) was paired with SSA to forecast stock prices [38]. A hybrid method based on SVM and EMD was proposed to improve oil price forecast accuracy [2]. Some research concentrated on the combination of artificial ANN with WT [17], EMD [43] and VMD [20]. Most, if not all, of these studies focused on one pair of forecasting technique and decomposition-reconstruction technique. None of them investigated the collective power of these techniques for crude oil price forecasting.

Forecast combination is a well-known approach to improving the performance of individual forecasts [4,26]. Many combination methods have been presented over the past decades, such as simple average, Bayesian methods, expert aggregation, and ensemble averaging using boosting, bagging or random forests [27]. The literature has shown that some simple combination methods, such as simple averaging, trimmed mean, and median are stable, promising and often superior to individual forecasts [39,49]. Since each individual forecast contributes to the combination, researchers have also looked into selection of individual forecasts before combining them. A Max-Linear-Relevance and Min-Linear-Redundancy based selection algorithm that provides a theoretical approach for the optimal sub-model selection [9]. An optimal model subset selection algorithm based on information theory was proposed in [31].

Another branch of forecast combination research looks into the relationship among the individual forecasts, from which the combination methods can be further divided into homogeneous and heterogeneous methods [32]. Homogeneous methods use the same induction algorithm to produce the models for combination, while heterogeneous methods integrate various forecasts from different algorithms with the same or diverse inputs. Both methods have been introduced to crude oil price forecasting. For instance, a homogeneous approach to investigating oil price predictability with a large set of predictors with 36 variables was taken in [47], while three forecasting models, ARIMA, exponential smoothing and dynamic regression was combined to predict the crude oil spot price [5].

Both combination methods have their pros and cons. A homogeneous method is largely easy to implement, but the forecast accuracy tends to be constrained by the underlying forecasting technique. A heterogeneous method may be robust, but can be computationally intensive. In this paper, we propose a semi-heterogeneous approach to combining crude oil price forecasts. Specifically, we interact four decomposition methods (e.g., WT, EMD, SSA, and VMD) with four forecasting techniques (e.g., AR, ARIMA, SVR, and ANN) to produce 16 individual forecasts for combination. We present a model selection method based on Tabu Search. The proposed approach offers an overproduce-and-select way to generate diverse base predictors. We will demonstrate that the proposed semi-heterogeneous combinations outperform the individual forecasts and the traditional heterogeneous combinations on average.

The rest of the paper is organized in five sections: Section 2 presents the preliminaries of four multi-band decomposition methods and four popular forecasting techniques. Section 3 introduces the modeling framework of the semi-heterogeneous combination. Section 4 presents the computational experiment and result analysis. The paper is concluded in Section 5.