

Chapter 10

Data Mining Methods for Crude Oil Market Analysis and Forecast

Jue Wang

Chinese Academy of Sciences, China

Wei Xu

Renmin University, China

Xun Zhang

Chinese Academy of Sciences, China

Yejing Bao

Beijing University of Technology, China

Ye Pang

The People's Insurance Company (Group) of China, China

Shouyang Wang

Chinese Academy of Sciences, China

ABSTRACT

In this study, two data mining based models are proposed for crude oil price analysis and forecasting, one of which is a hybrid wavelet decomposition and support vector Machine (SVM) model and the other is an OECD petroleum inventory levels based wavelet neural network model (WNN). These models utilize support vector regression (SVR) and artificial neural network (ANN) technique for crude oil prediction and are made comparison with other forecasting models, respectively. Empirical results show that the proposed nonlinear models can improve the performance of oil price forecasting. The findings of this research are useful for private organizations and governmental agencies to take either preventive or corrective actions to reduce the impact of large fluctuation in crude oil markets, and demonstrate that the implications of data mining in public and private sectors and government agencies are promising for analyzing and predicting on the basis of data.

DOI: 10.4018/978-1-60566-906-9.ch010

INTRODUCTION

The need for both private organizations and government agencies to utilize data in public and private sector activities is increasing in recent years, including collecting and managing the data, analyzing and predicting on the basis of data. For example, the manager of an energy sector, in order to make the right decisions, must know the expectation of international crude oil price, energy supply and demand. Simultaneously, he should know the main factors affecting oil fluctuation, demand and the supply capacity. Both the strategic and operational decisions of an organization or government agencies require the exploration of the current relationship among all the factors and construction of forecast models.

As a source of energy and chemical raw materials, crude oil plays an important role in the development of world economy. In recent years, the fluctuation of crude oil price becomes larger and larger, which not only directly affect global economic activities, but also bring risk to the oil-related enterprises and investors. Crude oil price is emerging as one of the hottest topics in the world. Influenced by many complicated factors, however, oil prices appear highly nonlinear and even chaotic (Panas and Ninni, 2000; Adrangi et al., 2001), which makes it rather difficult to forecast the future oil prices.

Although oil price forecasting is very difficult, it has fascinated many academic researchers and business practitioners in the past few decades. There have been substantial literatures on analysis and forecast of crude oil prices including qualitative and quantitative methods, on the basis of which many decisions with regard to oil prices have to be made (Fan et al. 2008). Among the qualitative methods, Nelson et al. (1994) used the Delphi method to predict oil prices for the California Energy Commission. Abramson and Finizza (1991) used belief networks, a class of knowledge-based models, to forecast crude oil prices.

Besides these qualitative methods, a large number of quantitative methods and models are developed to analyze and forecast crude oil prices. According to Zhang et al. (2008), the quantitative methods can be grouped into two categories: structure models and data-driven methods. Standard structure models outline the world oil market and then analyze the oil price volatility in terms of a supply-demand equilibrium schedule (Zhang et al. 2008). For example, Bacon (1991) discussed the factors determining the demand of oil, the supply of oil by OPEC and non-OPEC countries, and gave the forecast of crude oil prices. Al Faris (1991) analyzed the determinants of crude oil price adjustment in the world petroleum market. Data-driven methods include various models and approaches, such as traditional time series methods, econometric models and data mining techniques.

There are abundant studies on crude oil price prediction using time series and econometric methods. Huntington (1994) applied a sophisticated econometric model to predict crude oil prices in the 1980s. Abramson and Finizza (1995) utilized a probabilistic model for predicting oil prices. Gulen (1998) used co-integration analysis to predict the West Texas Intermediate (WTI) price. Barone-Adesi et al. (1998) suggested a semi-parametric approach for oil price prediction. Similarly, Morana (2001) offered a semi-parametric method for short-term oil price forecasting based on the GARCH properties of crude oil price. In a more recent study by Ye et al. (2002, 2005 and 2006), some short-term forecasting models of monthly WTI crude oil spot prices using OECD petroleum inventory levels are proposed. Lanza et al. (2005) investigated crude oil and oil products' prices using error correction models (ECM). Sadorsky (2006) used several different univariate and multivariate models such as TGARCH and GARCH to estimate forecasts of daily volatility in petroleum futures price returns.

As mentioned in Yu et al. (2008), the traditional time series and econometric models can provide

18 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/data-mining-methods-crude-oil/44289

Related Content

Graph-Based Modelling of Concurrent Sequential Patterns

Jing Lu, Weiru Chen and Malcolm Keech (2010). *International Journal of Data Warehousing and Mining* (pp. 41-58).

www.irma-international.org/article/graph-based-modelling-concurrent-sequential/42151

User-Centric Similarity and Proximity Measures for Spatial Personalization

Yanwu Yang, Christophe Claramunt, Marie-Aude Aufaure and Wensheng Zhang (2012). *Exploring Advances in Interdisciplinary Data Mining and Analytics: New Trends* (pp. 128-146).

www.irma-international.org/chapter/user-centric-similarity-proximity-measures/61172

Coronavirus Pandemic (COVID-19): Emotional Toll Analysis on Twitter

Jalal S. Alowibdi, Abdulrahman A. Alshdadi, Ali Daud, Mohamed M. Dessouky and Essa Ali Alhazmi (2022). *Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines* (pp. 1761-1782).

www.irma-international.org/chapter/coronavirus-pandemic-covid-19/308574

Conceptual Model and Design of Semantic Trajectory Data Warehouse

Michael Mireku Kwakye (2020). *International Journal of Data Warehousing and Mining* (pp. 108-131).

www.irma-international.org/article/conceptual-model-and-design-of-semantic-trajectory-data-warehouse/256165

Patient Oriented Readability Assessment for Heart Disease Healthcare Documents

Hui-Huang Hsu, Yu-Sheng Chen, Chuan-Jie Lin and Tun-Wen Pai (2020). *International Journal of Data Warehousing and Mining* (pp. 63-72).

www.irma-international.org/article/patient-oriented-readability-assessment-for-heart-disease-healthcare-documents/243414