Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 61 (2014) 509 – 512

Energy Procedia

The 6th International Conference on Applied Energy – ICAE2014

The Drivers of Oil Prices – A MI^3 Algorithm approach

Neha Sehgal[a], Krishan K Pandey[b,*]

[a]Jindal Global Business School, O. P. Jindal Global University, Sonipat, Haryana-131001, India
[b]College of Management & Economic Studies, University of Petroleum & Energy Studies, Dehradun-248007, India

Abstract

The debate on significance of numerous political, economic and financial indicators driving crude oil prices is perpetual. There is no single indicator which can provide a complete picture of how prices can be determined. Nor a simple combination of input indicators can provide accurate and robust price forecast methods. In particular, feature selection plays a key role in designing a forecasting model for oil prices. However, all existing method of predicting oil prices have accounted for non-linearity, non-stationarity and time-varying structure of crude oil prices but seldom focus on selecting significant features with high predicting power. Besides, there is lack of competent feature selection techniques based on associations and dependency of indicators for designing the input vector of oil price forecast. For this purpose, a novel two-stage feature selection method "MI^{3}Algorithm" is proposed for inferring non-linear dependence between oil prices and strategic indicators driving them by employing interaction information and mutual information as measure of redundancy(or synergy) and relevance. The study targets to figure out the importance and impacting mechanism of key indicators driving crude oil prices based on the proposed feature selection algorithm employing multi-layered perceptron neural network (MLP), general regression neural network (GRNN) and cascaded neural network (CNN) as forecasting engine for oil price prediction. The results confirmed the superiority of proposed algorithm compared to some other methods. Besides its high accuracy, the proposed algorithm provides non-redundant and most relevant features as compared to other methods employed in study.

Peer-review under responsibility of the Organizing Committee of ICAE2014
Keywords: Feature Selection ;Mutual Information; Interaction Information; Neural Networks; Oil price forecasting

1. Introduction

The "price of oil" is a critical factor with substantial impact on world economics, be they are part of OPEC or Non-OPEC countries. According to BP , oil remains the world's primary fuel, accounting to 33.1% of global energy consumptions. Oil prices have been steadily rising for several years and in July 2008 stand at a record high of above US$140/bbl. Later, it declined due to global economic crisis at the end of 2008, which took long to recover by 2010 as US$75/bbl. This rise or decline in oil prices

* Corresponding author. Tel.:+91-9458314387; +91-9760027312.
E-mail address: krishan.pandey@gmail.com.

Peer-review under responsibility of the Organizing Committee of ICAE2014
doi:10.1016/j.egypro.2014.11.1160

stimulates for studying in detail the factors behind movements in the "price of oil". There are large number of factors, which are complex, noisy, and uncertain influencing crude oil prices [1]. In order to build an effective model, careful attention should be paid on selecting informative and influential inputs which cause changes in prices [2]. However, until recently, the input variables of oil price forecast is selected on judgmental criteria or trial and error procedures [3]. Most of the studies were concentrated on non-linearity, non-stationarity and time varying properties of oil prices but seldom focus on designing a robust feature selection method for selecting significant inputs to improve forecasting accuracy. In this context, the study proposes a two stage non-linear feature selection method as discussed in next section.

2. MI3 Algorithm for feature selection

Inspired by the superiority of information-theoretic approaches, the study propose "MI3 Algorithm" composed of mutual information and interaction information for finding drivers of oil prices. The proposed algorithm consists of two stages. In the first stage, mutual information based irrelevance filter is used to selects the most relevant features from the set of candidate inputs. In the second stage, interaction information based redundancy filter removes the redundant features from selected relevant candidates. The selected features are used as set of input variables to build MLP, GRNN and CNN forecasting models. The structure of the proposed MI3 Algorithm is shown in fig 1.

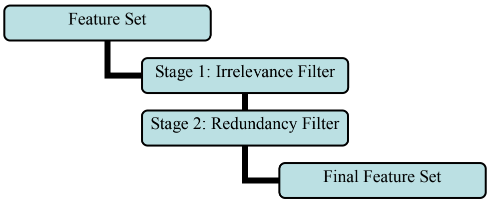

Fig 1 Structure of the proposed MI^3 feature selection algorithm

3. Numerical Results

Globalization hypothesis of oil prices moving together holds independently of whether the market is crashing or booming. In this study, WTI crude oil spot price is chosen as target variable as it is considered benchmark oil in global terms. The factors driving oil prices are classified into eight classes: Supply, Demand, Reserves, Inventory, Future Market, Exchange Rate, Economy and Weather. The data is collected on monthly basis from EIA, World Bank and Bloomberg databases from January 1994 – December 2011. The data is pre-processed by scaling to (-1, 1) as input for MLP and it is standardized (by subtracting the median and dividing by the interquartile range) as input for GRNN. The proposed algorithm is used to identify the set of most relevant and non-redundant features. In stage 1, mutual information is computed. The candidate inputs with the relevance rank (normalized relevance value with respect to maximum MI) and feature number are shown in table 1.

Table 1 Selected input by Stage 1 irrelevance filter for WTI crude oil price market

Feature (Rank, No.)	NYMEX future prices (1, 18)	India Consumption (2, 6)	OPEC Reserves (3, 10)	China GDP (4, 22)	India GDP(5, 23)
Feature (Rank, No.)	OECD Reserves(6, 9)	U.S GDP (7,26)	Reserve-Production Ratio (8, 12)	U.S. Inflation (9,24)	U.S. Refinery Capacity (10, 17)

Feature (Rank, No.)	Strategic Petroleum Reserves (11, 13)	Non-OECD Consumption (12, 4)	China Consumption (13, 5)	OPEC Supply (14, 2)	Non-OPEC Production (15, 1)
Feature (Rank, No.)	EUR/ USD (16, 21)	China Reserves (17, 11)	U.S. import from Non-OPEC (18, 15)	OECD Stocks (19, 14)	JPY/ USD (20, 20)
Feature (Rank, No.)	OPEC Spare Capacity (21, 8)	GBP/ USD (22, 19)	U.S. import from OPEC (23, 16)	OECD Consumption (24, 3)	Primary Energy Consumption (25, 7)
Feature (Rank, No.)	U.S. Cooling degree days (26, 28)	U.S. Heating degree days (27, 27)	Geopolitical and Economic Events (28, 25)		

Further, the 3-variable interaction information is computed. The algorithm in stage 2 start with maximum relevance rank variable X_{18}. The interaction information is negative for a single set $\{Y, X_{18}, X_{25}\}$. The redundant variable is filtered by comparing mutual information $I\{Y, X_{18})$ and $I\{Y, X_{25})$. The results obtained in table 1 shows $I\{Y, X_{18}) > I\{Y, X_{25})$, therefore, X_{25} is filtered out and the process continues for next subsequent ranked variables. The stage 2 of proposed algorithm has reduced the number of candidate input to 16 from 28 (approx. 50% of actual number) which are non-redundant and relevant in nature. In this experiment, the performance of two stage feature selection MI^3 algorithm is evaluated with known feature selection techniques as: Modified Relief + Mutual Information (MR + MI) [4], Modified Relief (MR) and Correlation feature selection (CFS). The proposed MI^3 algorithm with three forecasting engines is compared with 6 single-stage models (MR / CFS) and 3 two-stage models (MR + MI). The ultimate goal of this study is to select informative inputs to build effective forecasting model for crude oil price. The performance criteria to compare the effectiveness of the different models are root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) and are shown in table 2 for 12 ensemble models. MI^3 algorithm with GRNN and MLP has performed superior to other competing ensemble methods (least RMSE, MAE & MAPE). CFS seems to be performed well with MLP but there are several limitations regarding relevance, redundancy and basic assumption of conditional independence associated with it. The proposed algorithm is superior in finding the set of non-redundant features compared to features selected through (MR+MI) algorithm as there exist two pairs of features selected for which interaction information is negative. The proposed algorithm is a fully automatic algorithm that doesn't require user to specify the number of features to be extracted or to specify any threshold. Most of the competing feature selection method fails to perform well as they assume that features are conditional independent within the classes. The proposed algorithm overcomes this limitation using the concept of interaction information. The explanatory power for oil price using sixteen selected features is 99.01%, indicating that the variable reduction is reasonable and will have no essential influence on subsequent analysis. The final features selected from proposed algorithm are shown in table 3. It shows enormous impact of emerging economies in driving crude oil prices. Our results have proved the importance of Non-OECD consumption through recent change in data post 2009. Further, Non-OPEC production which constitutes production from developed countries is also influential in deciding direction of crude oil prices. Results indicate that NYMEX future prices are not a sole indicator driving the directions of spot oil prices.

Table 2 Performance evaluation for WTI crude oil price forecasting

Model	RMSE	MAE	MAPE	Model	RMSE	MAE	MAPE
MI^3 + GRNN	0.32	0.25	0.92	MR+GRNN	2.38	1.55	3.99
MI^3+ MLP	0.60	0.46	1.68	MR+MLP	1.27	0.85	2.81
MI^3+ CNN	2.27	1.73	5.07	MR+CNN	4.11	2.81	8.65

MR+MI+ GRNN	1.51	0.96	2.38	CFS+GRNN	1.82	1.15	2.94
MR+MI+MLP	0.48	0.29	0.89	CFS+MLP	0.39	0.29	0.92
MR+MI+CNN	1.46	0.97	2.67	CFS+CNN	3.01	2.21	7.11

Table 3 Explanatory power of 16 selected factors based on MI^3 algorithm

Feature	Non-OPEC Production	OPEC Supply	Non-OECD Consumption	China Consumption	India Consumption	OECD Reserves	OPEC Reserves	China Reserves
%	1.21	0.01	1.96	1.63	3.73	0.16	9.49	5.73
Feature	Reserve Production Ratio	Strategic Petroleum Reserve	U.S Refinery Capacity	NYEMX Future Price	China GDP	India GDP	U.S Inflation	U.S.GDP
%	15.09	2.96	4.27	15.89	10.67	5.32	11.22	10.67

4. Conclusions

In this paper, we presented a novel two stage MI^3 algorithm for selecting informative inputs to build a forecasting model of crude oil forecasting. The algorithm is composed of informatics-theory approaches for studying the dependency and association of input variables for prediction problems. The results shows superiority of proposed algorithm in finding most relevant and non-redundant inputs compared to some known methods. The results highlighted MI^3 + GRNN as best forecasting tool for crude oil prices. The proposed algorithm reduced the number of candidates to approximately half of actual number of inputs.

5. Reference

[1]. De Souza e Silva, E. G., Legey, L. F., & de Souza e Silva, E. A. (2010, November). Forecasting oil price trends using wavelets and hidden Markov models. Energy Economics, 32(6), 1507-1519.

[2]. Haidar, I., Kulkarni, S., & Pan, H. (2008). Forecasting Model for Crude Oil prices Based on Artificial Neural Networks. International Conference on Intelligent Sensors, Sensor networks and Information Processing, 103-108. IEEE.

[3]. Jammazi, R. & Aloui, C. (2012). Crude oil price forecasting: Experimental evidence from wavelet decomposition and neural networks. Energy Economics, 34, 828-841

[4] Amjady, N. & Daraeepour, A. (2009). Design of input vector for day-ahead price forecasting of electricity markets. Experts Systems with Applications, 36(10), 12281-12294.

Biography

Neha Sehgal is a Ph.D. candidate at College of Management & Economic Studies, University of Petroleum & Energy Studies. She is also the Research Associate at Jindal Global Business School, O. P. Jindal Global University. Her main research interests include data mining, energy economics and time series econometrics.

Krishan K. Pandey is Assistant Dean (Research) and Associate Professor at College of Management & Economic Studies, University of Petroleum & Energy Studies. His main research interests include statistical inference, statistical modeling for bio-fuels, LCA and EIA for energy crops and time series econometrics.