Aftermath of 2008 Financial Crisis on Oil Prices

Neha Sehgal[1] and Krishan K. Pandey[2]
[1]*Jindal Global Business School, O. P. Jindal Global University, Sonipat, 131001, Haryana, India*
[2]*College of Management & Economic Studies, University of Petroleum & Energy Studies, 248007, Dehradun, India*

Keywords: Feature Selection, Mutual Information, Interaction Information, Neural Networks, Oil Price Forecasting.

Abstract: Geopolitical and economic events had strong impact on crude oil markets for over 40 years. Oil prices steadily rose for several years and in July 2008 stood at a record high of $145 per barrel. Further, it plunged to $43 per barrel by end of 2008. There is need to identify appropriate features (factors) explaining the characteristics of oil markets during booming and downturn period. Feature selection can help in identifying the most informative and influential input variables before and after financial crisis. The study used an extended version of MI^3 algorithm i.e. I^2MI^2 algorithm together with general regression neural network as forecasting engine to examine the explanatory power of selected features and their contribution in driving oil prices. The study used features selected from proposed methodology for one-month ahead and twelve-month ahead forecast horizon. The forecast from the proposed methodology outperformed in comparison to EIA's STEO estimates. Results shows that reserves and speculations were main players before the crisis and the overall mechanism was broken due to 2008 global financial crisis. The contribution of emerging economy (China) emerged as important variable in explaining the directions of oil prices. EPPI and CPI remain the building blocks before and after crisis while influence of Non-OECD consumption rises after the crisis.

1 INTRODUCTION

Oil prices are dependent on numerous indicators but there influence is subject to happening of geopolitical and economic events. Oil prices steadily rose for several years post 9/11 attacks and in July 2008 stood at a record high of $145 per barrel due to low spare capacity. Further, due to global financial crisis of 2008, oil prices plunged to around $43 per barrel by end of 2008. In quarter 1 of 2009, OPEC slashed production targets by 4.2 mmbpd and thus oil prices rose from $43 per barrel to $91 per barrel by end of 2011. The question that arises is whether this rise or decline in oil price is entirely due to shift in demand-supply framework or are there any other political or economic indicators to blame? And if there are other significant indicators driving oil prices, how does the explanatory power and contribution of factors driving oil prices changes during booming and downturn period. A study by Bhar and Malliaris (2011) concluded that price increases during financial crisis of 2007-2009 were so substantial that additional factors other than demand and supply were needed to explain such drastic shifts. Another study (Fan and Xu, 2011) used break test to divide the price fluctuations in oil markets after 2000 into three stages: January 2000-March 2004, March 2004-June 2008 and June 2008-September 2009. Their study has shown that in different time periods, the main drivers of oil prices changed and their direction and degree of influence will change over time.

There is colossal collection of data for factors, ranging from demand-supply, inventories, reserves to varied market, is enormous and dynamic. An important task is to discover knowledge by identifying useful patterns (most influential and informative set of factors driving oil prices) in data. Till date, researchers employing structural or financial models for predicting oil prices have accounted for non-linearity, non-stationary or time-varying structure of the oil prices but seldom have focused on selecting significant features with high prediction power. Most of the researchers have considered predictor variables for oil price prediction based on judgmental criterion or trial and error method. Little attention is paid on selecting most influential and informative factors and more on assessing new techniques for oil price forecasting. Therefore, feature selection plays an important role in forecasting oil prices. An appropriate set of

Sehgal, N. and Pandey, K..
Aftermath of 2008 Financial Crisis on Oil Prices.
In *Proceedings of the 7th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2015) - Volume 1: KDIR*, pages 235-240
ISBN: 978-989-758-158-8
Copyright © 2015 by SCITEPRESS – Science and Technology Publications, Lda. All rights reserved

features can help in high prediction performance and thus, due care should be taken to select a set of relevant and non-redundant features However, conventional feature selection methods require number of features to be extracted or a strict assumption of conditional independence, and still couldn't provide the minimal set of features that are most relevant and non-redundant for the study. The basic assumption of conditional independence of feature selection methods degrades the performance of model if features are strongly inter-connected. Most of the real world problems contain features that are strongly inter-related to each other. Due to above mentioned research gaps; there is lack of robust feature selection method to select relevant and non-redundant factors for oil price forecasting which can incorporate complexities of crude oil prices. Hence, to overcome the limitations of existing pool of methods, this study used I^2MI^2 feature selection algorithm when features are strongly dependent on each other and are non-linear.

2 I^2MI^2 ALGORITHM FOR FEATURE SELECTION

The novel three stage feature selection method called I^2MI^2 algorithm is an extended version of MI^3 Algorithm (Sehgal and Pandey, 2014) build on pillars of interaction information and mutual information. It is used for selecting relevant and non-redundant features that drive oil price. The proposed algorithm consists of three stages. In the first stage, mutual information is computed between target variable and candidate inputs. The variables are ranked based on normalized mutual information value and the irrelevant features are filtered out based on a threshold value. The selected variables are the list of irrelevant but redundant features. To overcome redundancy, in stage two, three-variable interaction information is computed among the selected features in stage one. The set of selected features having negative interaction information are used to filter out the redundant features.

The study incorporates the concept of interaction information so as to filter redundant input variables instead of correlation analysis or partial correlation analysis. Interaction information is favoured over correlation analysis as it measures non-linear dependency. This stage provides list of features that are relevant and non-redundant in nature. Further, in the third stage, mutual information is computed between the selected features from stage two and ranked according to normalized mutual information value. Depending on a threshold value, redundant features in stage three are filtered according to relevance rank in stage one. The selected features are used to build neural networks for oil price prediction. The performance of proposed feature selection algorithm is compared with Correlation based Feature Selection (CFS), Modified Relief (MR) and Modified Relief + Mutual Information (MR + MI) (Amjady and Daraeepour, 2009) feature selection methods. The performance criterions used for comparing I^2MI^2 algorithm with other algorithms are RMSE, MAE and MAPE.

The proposed algorithm I^2MI^2 with GRNN as forecasting engine has performed the best among all other feature selection methods. I^2MI^2 algorithm has lowest RMSE, MAE and MAPE as 1.29, 0.96 and 2.51 respectively. The reason for the best performance lies in the fact that the final selected features from proposed algorithm are 100% non-redundant and relevant for the study. Two stage (MR + MI) with CNN as forecasting engine as proposed by Amjady and Daraeepour (Amjady and Daraeepour, 2009) has not performed better than proposed algorithm. I^2MI^2 algorithm is fully automatic algorithm and doesn't require user to specify the number of features to be selected. I^2MI^2 algorithm can provide the minimal representative set of features for regression problems in business, biostatistics, applied energy and many more disciplines.

3 NUMERICAL RESULTS

For analysing the different mechanism in the falling and rising period of oil prices, two sub-periods are considered: January 2004-July 2008 and August 2008-December 2012, before and after 2008 financial crisis, respectively. The data collected for factors driving oil prices are classified into eight major classes: Speculations (2), Supply (3-4), Demand (5-8), Reserves (9-15), Inventory (16-18), Exchange Market (19-22), Stock Market (23) and Economy (24-26) as shown in Table 1. The features are selected on the basis of extensive literature review. For each sub-period, I^2MI^2 algorithm is applied to select minimal set of relevant and non-redundant factors that leads to high prediction performance for oil prices. General Regression Neural Network model is used as forecasting engines to analyse the explanatory power of selected features and their contribution in driving oil prices. The proposed methodology is used to forecast the new characteristics of oil prices one-month and twelve-month ahead before and after the crisis. The forecasts from the proposed methodology are compared with EIA's STEO January 2013 onwards forecast reports.

3.1 Sub-Period 1: January 2004-July 2008

The goal of stage one is to provide relevant features based on mutual information irrelevance filter. The step by step procedures followed in stage 1 of proposed I^2MI^2 algorithm are as follows. The candidate features (column 1) with the relevance rank (column 2) and their normalized relevance rank value (column 3) with the respect to maximum mutual information with oil prices are shown in Table 1. Column 4 provides the feature number. Based on a low threshold value Th1, feature number 16, 5, 3 and 15 can be filtered out by relevance filter. The goal of stage two is to provide non-redundant and relevant features based on redundancy filter. The three-variable interaction information between target variable and features selected from stage 1 is computed. Since interaction information $I(Y, X_i, X_j)$ is a symmetric measure; it cannot derive the direction whether X_j inhibits the correlation between (Y, X_i) or X_i inhibits the correlation between (Y, X_j). Therefore, it become difficult to filter the redundant variable from the set of relevant features (X_i, X_j) when interaction information is negative. In this thesis, this limitation of interaction information is relieved by focusing on mutual information between target and input variables $I(Y, X_i)$. The algorithm in stage two starts with maximum relevance rank variable from stage one. The variable EPPI(26) is ranked first as evident from Table 1. Add X_{26} to set S_2. For the first relevance ranked variable X_{26} there are seven set $\{Y, X_{26}, X_j\}$ where j = {3, 4, 5, 8, 16, 17, 21} for which interaction information is negative. The question that arises here is whether X_i inhibits the correlation between Y and X_{26} or X_{26} inhibits the correlation between Y and X_i. The redundant variable is filtered by comparing mutual information $I(Y, X_{26}\}$ with $I(Y, X_j)$ for each j. The results thus obtained in Table 1 shows that mutual information $I(Y, X_{26}) > I(Y, X_j)$ for each j. Therefore, the variables X_j for j = {3, 4, 5, 8, 16, 17, 21} are redundant variables and must be filtered out from the list of relevant and non-redundant variables. Similarly, the process holds for next ranked variable X_{25} from Table 1. The features thus selected through stage two are shown in Table 2. The numbers of candidate inputs (N) are reduced from 25 to 11 in stage two; i.e. to less than 50% of the actual number of input variables. The algorithm in stage three starts with maximum relevance rank variable X_{26} from Table 1. By default, X_{26} is considered as part of final set. Now, consider the next relevance rank feature X_{25}.

According to the pre-specified threshold value Th2, variables from stage two are filtered out based on mutual information between features. Since mutual information $I(X_{26}, X_{25}) > Th2$, therefore, X_{25} is filtered out by redundancy filter. The final sentence of a caption must end with a period.

Table 1: Relevance rank based on stage one of proposed algorithm.

Feature	Rank, No.
EPPI (Producer price index)	1, 26
CPI (Consumer price index)	2, 25
NCPP (Speculations)	4, 2
GDP (U.S Gross domestic product)	5, 24
SPR (Strategic Petroleum Reserve)	6, 12
GU (GBP/USD)	7, 20
Non-OECD-C (Non-OECD consumption)	8, 7
EU (EUR/USD)	9, 22
DER (U.S. Dollar Exchange rate)	10, 19
RP (Reserve Production Ratio)	11, 11
OPEC-R (OPEC Reserves)	12, 14
RC (U.S. Refinery Capacity)	13, 18
OECD-R (OECD Reserves)	14, 13
OPS (OECD Petroleum stocks)	15, 10
CC (China consumption)	16, 6
OSC (OPEC Spare capacity)	17, 9
OPEC-S (OPEC Supply)	18, 4
IC (India Consumption)	19, 8
JU (JPY/USD)	20, 21
I-Non-OPEC (Petroleum Import from Non-OPEC)	21, 17
I-OPEC (Petroleum Import from OPEC)	22, 16
OECD-C (OECD Consumption)	23, 5
Non-OPEC-P (Non-OPEC Production)	24, 3
CR (China Reserves)	25, 15

For the next relevant ranked feature X_m, calculate maximum mutual information Max(MI) between X_m and previously selected candidates in set stage three by redundancy filter. If Max(MI) > Th2 for any set, then X_m is filter out by redundancy filter. Otherwise, X_m is added to the final selected features set. The algorithm will run iteratively for all 11 selected variables from stage two. The final selected features from the proposed I^2MI^2 algorithm are EPPI (26),

Table 2: Filtered features by redundancy filter in stage two.

Filtered Feature (Stage 2)	No., Rank
EPPI	26, 1
CPI	25, 2
DJI	23, 3
NCPP	2, 4
GDP	24, 5
SPR	12, 6
Non-OECD-C	7, 8
DER	19, 10
RP	11, 11
OPEC-R	14, 12
OECD-R	13, 14

NCPP (2), SPR (12), DER (19) and RP(11). Thus, five out of twenty five variables were selected to represent fluctuations in oil prices before the crisis. The selected features are used as input variables to General Regression neural networks forecasting engine. The performance of proposed feature selection algorithm with GRNN forecasting engine is evaluated based on RMSE, MAE and MAPE. The proposed ensemble model is used to forecast in-sample and out-of-sample. Firstly, in order to compare the model's capability with other models, nearly 4.4-year (January 2004-July 2008) monthly data is used for training and validation. In-sample evaluations are shown in Table 3. The model is used to produce one and twelve-month ahead out-of-sample forecasts from August 2008 till July 2009. To evaluate the performance of our model, we compare it with forecasts shown in EIA's STEO reports from August 2008 onwards. Out-of sample evaluations are shown in Table 4. The proposed methodology performed better in terms of MAE for one-month ahead forecasts as compared to EIA's STEO forecasts but not in terms on RMSE and MAPE. It is evident from Table 4 that the proposed model performed superior as compared to STEO model for twelve-month ahead forecasts during extreme complex and volatility phase of oil prices. It also shows that the model does very well based on input variables selected by proposed algorithm as compared to EIA's STEO forecasts. The proposed methodology performed more accurately in long-run forecasting as compared to short-run when the market is too complex and highly volatile. The explanatory power for oil prices using five selected features is 97.6% before the crisis, indicating that the variable reduction is reasonable and that it will have no essential influence on subsequent analysis.

Table 3: In-sample performance of proposed methodology.

Proposed Methodolgy	
RMSE	3.55
MAE	2.74
MAPE	4.13

Table 4: Out-of-Sample forecast comparison.

Model	RMSE, MAE, MAPE
One-Month (Proposed)	8.24, 9.74, 13.27
One-Month(STEO)	6.85, 9.91, 10.82
Twelve-Month(Proposed)	31.9, 34.85, 63.3
Twelve-Month(STEO)	67.59, 62.49, 122.81

3.2 Sub-period 2: August 2008 - November 2012

The proposed methodology is used to find most influential and informative features using the same methodology as discussed in section 3.2. The tables corresponding to stage one (Table 5) and stage two (Table 6) are shown for references are shown in Appendix A. The final set of features of features selected in this subgroup are EPPI(26), DJI(23), CC(6) and CR(15). In-sample performance of proposed methodology in this sub-period is shown in Table 7. The results from Table 8 show superior performance of our proposed model in comparison to EIA's STEO model for both one-month and twelve-month ahead forecasts. The MAPE for the whole period (December 2012-November 2013) is 6.27 while RMSE and MAE are 6.47 and 6.30 respectively for twelve-month ahead time period. Similarly, the MAPE is 2.12 while RMSE and MAE are 2.64 and 2.01 for one-month ahead forecast horizon. Our model performed well in both in-sample and out-of-sample forecast horizons. The explanatory power of oil prices using four selected features is 93.8% after the crisis, indicating that the variable reduction is reasonable and that it will have no essential influence on subsequent analysis.

4 CONCLUSIONS

The detail regarding the factors contribution to oil prices before and after 2008 financial crisis is as follows. The importance of 11 variables (OPEC-S, Non-OPEC-P, CC, Non-OECD-C, IC, OSC, OECD-R, OPEC-R, CR, RC, JU) increases, 10 variables (NCPP, Non-OECD-C, OPS, RP, SPR, I-OPEC, I-Non-OPEC, DER, GU, EU, GDP) decreases and for

4 variables (EPPI, CPI, DJI, OECD-C) remain unchanged. The analysis reveals that various driving factors show some new characteristics after the financial crisis. Same is discussed as follows:

- EPPI and CPI have taken up first two positions before and after crisis. Speculation position has declined significantly after crisis due to high fluctuation in oil prices.
- Influence of Non-OECD consumption has increased after crisis but OECD consumption remains at same pace.
- The explanatory powers of China consumption and China reserves have increases and they both have emerged as important variables driving oil prices.
- The explanatory power of strategic petroleum reserves and reserve-production ratios have weaken after crisis.
- Global economic recession weaken US dollar together with GU and EU exchange market. On the other hand, JU exchange market power increased post crisis.
- The explanatory power of imports from OPEC declined whereas import from Non-OPEC increased. Due to disturbance in oil market as OPEC cuts target production, U.S is heading for sustainable solutions.

Overall, before the crisis, NCPP, EPPI, DER, SPR and RP were the major players that influence oil prices volatility. Before the crisis, DER was the major factor boosting change in oil prices together with RP. SPR played a major role in influencing oil prices due to disturbance created by cuts in OPEC production or OPEC news. On the contrary, the original mechanism of crude oil market was destroyed by 2008 financial crisis and the relationship of EPPI and DER with oil prices strengthened after crisis. China consumption and its reserves emerged as important influencing variables in recent times. The supply-demand framework has weaken after crisis and the influence of emerging economies has increased.

REFERENCES

Bhar, R., & Malliaris, A. G. 2011. Oil prices and the impact of the financial crisis of 2007–2009. Energy Economics, 33(6), 1049-1054.

Fan, Y., & Xu, J. H. 2011. What has driven oil prices since 2000? A structural change perspective. Energy Economics, 33(6), 1082-1094.

Sehgal, N., & Pandey, K. K. 2014. The Drivers of Oil Prices–A MI 3 Algorithm Approach. Energy Procedia, 61, 509-512.

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

APPENDIX

Table 5: Relevance rank based on stage one of proposed algorithm.

Feature	Rank, No.
EPPI	1, 26
CPI	2, 25
DJI	3, 23
CC	4, 6
Non-OPEC-C	5, 7
GDP	6, 24
IC	7, 8
OECD-R	8, 13
OPEC-S	9, 4
SPR	10, 12
OPEC-R	11, 14
RC	12, 18
OSC	13, 9
JU	14, 21
RP	15, 11
EU	16, 22
NCPP	17, 2
Non-OPEC-P	18, 3
I-Non-OPEC	19, 17
OPS	20, 10
DER	21, 19
GU	22, 20
OECD-C	23, 5
CR	24, 15
I-OPEC	25, 16

Table 6: Filtered features by redundancy filter in stage two.

Filtered Features(Stage 2)	No., Rank
EPPI	26, 1
CPI	25, 2
DJI	23, 3
CC	6, 4
OECD-R	13, 8
SPR	12, 10
OPEC-R	14, 11
RP	11, 15
CR	15, 24

Table 7: In-sample performance of proposed methodology.

Proposed Methodology	
RMSE	4.41
MAE	3.41
MAPE	4.31

Table 8: Out-of-Sample forecast comparison.

Model	RMSE, MAE, MAPE
One-Month (Proposed)	2.64, 2.01, 2.12
One-Month(STEO)	2.86, 3.51, 2.9
Twelve-Month(Proposed)	6.47, 6.3, 6.27
Twelve-Month(STEO)	9.81, 8.36, 8.31