INTERNATIONAL JOURNAL OF RESEARCH IN BUSINESS AND SOCIAL SCIENCE 9(7)(2020) 123-129

Business & Social Science
IJRBS

Research in Business & Social Science

IJRBS VOL 9 NO 7 SPECIAL ISSUE ISSN: 2147-4478

Available online at www.ssbfnet.com
Journal homepage: https://www.ssbfnet.com/ojs/index.php/ijrbs

Screening for light crude oil and market comovements

Omid Faseli

Institute of Information System Engineering, Vienna University of Technology, Vienna, Austria

Crossref

ARTICLE INFO

Article history:

Received 24 October 2020
Received in rev. form 14 Nov.. 2020
Accepted 18 November 2020

Keywords:

Data mining, pattern recognition, correlation cluster map, clustering, comovement, light crude oil, WTI, Brent

JEL Classification:
C8, C80, C82

ABSTRACT

This study aimed to perform a screening for economic interrelationships among market participants from the stock market, global stock indices, and commodities from fossil energy, agricultural, and the metals sector. Particular focus was put on the comovements of the light crude oil benchmarks West Texas Intermediate (WTI) and Brent crude oil. In finance research and the crude oil markets, identifying novel groupings and interactions is a fundamental requirement due to the extended impact of crude oil price fluctuations on economic growth and inflation. Thus, it is of high interest for investors to identify market players and interactions that appear sensitive to crude oil price volatility triggers. The price development of 14 stocks, 25 leading global indices, and 13 commodity prices, including WTI and Brent, were analyzed via data mining applying the hierarchical correlation cluster mapping technique. All price data comprised the period from January 2012 – December 2018 and were based on daily returns. The technique identifies and visualizes existing hierarchical clusters and correlation patterns emphasizing comovements that indicate positively correlated processes. The method successfully identified clustering patterns and a series of relevant and partly unexpected novel comovements in all investigated economic sectors. Although additional research is required to reveal the causative factors, the study offers an insight into in-depth market interrelationships.

Introduction

The generation and availability of reliable prediction models highly depend on finding relevant key players that impact crude oil prices. The search for correlation patterns in various economic fields is a first step and prepares the ground to investigate the underlying factors for these correlations. Methods for clustering and visualization analysis are supposed to be an important tool to shed light on the relationships between different market factors. It is assumed that market participants being highly correlated with crude oil price movements may also be vulnerable to high crude oil price fluctuations and volatility peaks. The crude oil comovers would then participate in these fluctuations during highly volatile oil price episodes. Special focus was put on the correlation partners of the light crude oil benchmarks WTI and Brent. Both hold a particular position in the global market since they show a signaling effect on investors and traders.

The study searches for crude oil price return comovements in the global markets and aims to reveal hidden groupings and correlations. It needs to be emphasized that comovements only indicate positively correlated price movements (Barberis et al., 2005). The comovers are important guideposts in the search for novel economic interactions. The research was conducted among 14 stocks of oil companies, 25 leading global indices, and 13 commodities from the energy, agricultural, and metals sector, applying the hierarchical correlation cluster mapping technique.

The agglomerative clustering method uses a variation of Ward's algorithm to identify the clusters and correlation patterns. All price data comprised the period from January 2012 – December 2018 and were based on daily returns. The technique can be applied as a first step screening tool for correlation partners placed within an expandable data mining approach that aims to predict price developments as the final target. Clustering techniques are well established and explored in data science and can be used for various applications in all research fields, including finance. Different approaches emerged to extract the naturally existing groups of related data objects. The variety of clustering algorithms follow the two basic approaches of agglomerative and divisive clustering. Agglomerative clustering represents the bottom-up approach, which is realized in Ward's method used for this project. Each data

* *Corresponding author. ORCID ID: 0000-0002-6713-1553*

https://doi.org/10.20525/ijrbs.v9i7.949

point starts with its own single cluster, which is then merged with similar data points on the next iterative level. By contrast, divisive clustering represents the top-down approach, which is used from the k-means algorithm dividing one cluster into two different clusters, and so on (Kilitcioglu, 2018; Pierson, 2015). Depending on the objective of a project, different algorithms can be beneficial. Techniques and tools to reveal natural groupings and clusters between data points are required in many fields. In finance research and the crude oil markets, identifying novel groupings and interactions is a fundamental requirement. Knowing the interaction points facilitates decision-making processes and supports the prediction of market behavior and price development. The method successfully identified clustering patterns and a series of relevant and partly unexpected novel comovements in all investigated economic sectors. Although additional research is required to investigate the causative factors, the study offers an insight into in-depth market interrelationships.

This research is structured as follows: the introduction in the first section presents an overview of the method and outlines the study's objective. The literature review provides related work referring to the crude oil market and the application of hierarchical correlation clustering to characterize current developments in both fields. It is followed by the research and methodology, characterizing the implemented method and algorithm and providing information about the data preprocessing steps. The publication is completed by the results and discussion section, presenting the visualized cluster mapping results. The project's outcome is summarized in the concluding section, reflecting current and possible future implementations of the screening. Areas for future investigations are mentioned. All tables are provided in the appendix to facilitate the reading flow. Acronyms and abbreviations are indicated in the appendix.

Literature Review

High levels of oil price fluctuations and spontaneous price movements are inseparably associated with increasing inflation and unstable economic growth, which can cause adverse effects on oil-importing and oil-exporting countries (Zhang et al., 2007; Kilian & Vega, 2011). The relationships between the crude oil market and a wide range of other market participants correlating with oil price development are of great interest for investors for whom most recent events and information generally the most relevant are (Baumohl, 2013). Crude oil price development takes place in economic surroundings influenced by various key players, among which other commodities, oil companies, or leading global indices are supposed to show comovements with crude oil. Volatility is a basic characteristic of dynamic financial markets and is used for risk assessment in asset pricing (Xiao & Aydemir, 2007). However, high levels of volatility were described to increase inflation [Libo et al., 2011] and to inhibit economic growth (Ebrahim et al., 2014). A wide range of factors was reported to show an impact on crude oil prices (E.I.A., 2019), among which spontaneous global events and geopolitical developments are listed (Kilian, 2010). Scheduled macroeconomic news announcements were found to correlate with crude oil price volatility (Faseli & Zamani, 2016; Faseli, 2019; 2020).

It was also demonstrated that high oil price fluctuations enhance uncertainty in commodity markets (Bakas & Triantafyllou, 2018; 2019), and the effects of uncertainty shocks were analyzed by (Su et al., 2018). Authors have intensively studied clustering methodologies over the decades since the second half of the last century. Ward (1963) laid the foundation for hierarchical agglomerative clustering. A variation of the original algorithm implemented in the programming language Python was applied in this project. Sun et al. (1996) emphasize the complexity of economic forecasting methods and analyzed Chinese stock price data using competitive learning neural networks (CLNNs). CLNNs adopt supervised learning and provide a two-stage cascaded clustering procedure capable of revealing hierarchical clusters in the data samples. The clustering method was found to be easily controlled and communicated and thus useful for stock market analysts. A different approach to clustering scenarios was performed by Ye et al. (2011), who proposed a combined hybrid methodology of k-means clustering algorithm in a self-organizing map network for a novel stock trading rule. They used 5min intraday data of Hushen 300 Index over January 5, 2009, to April 15, 2010, 70% of data were used as a training sample and 30% as a test sample. The performance of the trading rule based on hybrid clustering was compared with classical trading techniques. They revealed intraday patterns that can be used to support investment decisions. Noise variables and outliers in data samples affect data analysis processes, especially for real-world data (Filzmoser et al., 2008). Group structures can be masked by noise variables (Gordon, 1999; Becker & Gather, 1999), which should be removed. The selection of appropriate clustering parameters also plays an important role. Brodinová et al. (2019) proposed a k-means based clustering algorithm capable of grouping outlier and noise-containing data sets. The algorithm contains an automatic weighting function for each data point. Since k-means requires the pre-determination of cluster numbers, a framework for the pre-selection of cluster numbers was introduced.

Research and Methodology

This study's clustering method combines the benefits of hierarchical clustering, including branched dendrograms and visualized correlation patterns plotted as colored heat-maps. The data covered the observation period from January 2012 to December 2018, using daily returns. Clustering was performed using *scipy.cluster.hierarchy.ward* algorithm from programming language Python's SciKit library. Ward's method applies the Ward variance minimization algorithm, which is a variation of the original approach of Ward (1963). A range of variations was created of Ward's agglomerative clustering method using criteria for minimization and alternatively for maximization. Researchers modified the original method several times. (Murtagh & Legendre, 2014).

Ward's algorithm (Python) - New entry $d(u,v)$:

$$d(u,v) = \sqrt{\frac{|v|+|s|}{T}d(v,s)^2 + \frac{|v|+|t|}{T}d(v,t)^2 - \frac{|v|}{T}d(s,t)^2}$$

Let *d(u,v)* be the new entry, were the newly joined cluster *u* consists of the clusters *s* and *t*, *v* represents an unused cluster in the forest, $T=|v|+|s|+|t|$, $|*|$ denotes the cardinality of its argument. It is also described as the incremental algorithm (SciPyCommunity, 2020).

The clustering procedure commences with *n* clusters consisting of all single data points in the sample. These individual *n* clusters are merged by conducting iterative computational steps. Thus, the clusters are combined, whereby the output dendrogram height represents the distances between the clusters. The procedure finally combines all data points to a single cluster consisting of a varying number of sub-clusters. Each newly added cluster minimizes variance. At each step, the selection of a new cluster requires all possible combinations of clusters to be considered (Glen, 2018). Using *seaborn.clustermap* the matrix data set is plotted as a hierarchically clustered correlation heat-map. The cluster map visualizes the degree of correlation by using different colors and facilitates the intuitive understanding and interpretation of the results. Highly correlated areas are colored in light cream and orange, purple indicates low or close to zero correlation, and black fields denote minimal to no correlation. The correlation coefficients are indicated between -1 (negatively correlated items) and +1 (positively correlated items). Zero indicates no correlation.

Result and Discussion

Hierarchical correlation cluster mapping delivers results in the form of dendrograms indicating hierarchical clusters and cluster heat-maps visualizing the outcome of correlation cluster analysis. Additionally, the correlation coefficient values are provided. The dendrograms' top or root represents the entire data set, which is then subdivided into several branches or levels that are usually unequally sized. Visualization of the correlation patterns using color gradations in a correlation heat-map provides easily understandable results.

Energy Sector Stocks and Commodities (Figure 1):

The energy sector correlation cluster map in Figure 1 comprises daily price returns of 14 oil company stocks and fossil energy futures returns (WTI, Brent, heating oil, and natural gas, see Figure 1). The dendrogram indicates one dominant cluster comprising two subclusters and natural gas as a separated position that is not assignable to the available subclusters. The highest correlation pattern within the heat-map is indicated between the fossil energy futures WTI, Brent, and heating oil. Interestingly, natural gas futures were not grouped with the other applied fossil energy futures.

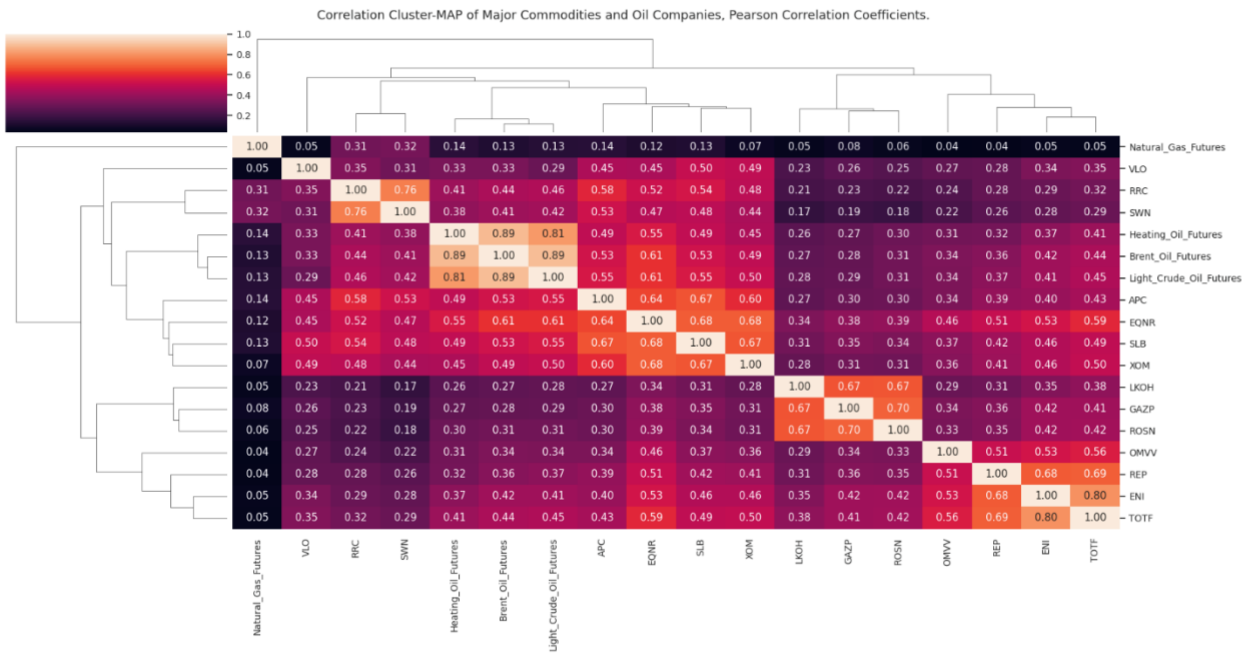

Figure 1: Correlation cluster map of daily energy, soft and hard commodity futures returns, observation period 2012-2018.

Gas futures occupy a separated position outside the clusterings pattern showing a low correlation with the United States (US) oil company stocks RRC and SWN. Other relationships were identified among the daily stock futures returns of APC, EQNR (Norway), SLB (Netherlands), and XOM (US), as well as among the Russian oil company stocks Lukoil, Gazprom, and Rosneft. The closely related price development of ENI (Italy), TOTF (France), REP (Spain), and OMVV (Austria) is denoted on the right side of the heat-map's bottom in decreasing order.

Energy, Soft and Hard Commodities (Figure 2)

In Figure 2, daily futures price returns were applied. The commodities were chosen from the fossil energy sector (WTI, Brent, heating oil, and natural gas), soft commodity or agricultural sector (sugar, coffee, corn, wheat), and from the metals sector (copper, platinum, gold, and silver). The observation period covered January 2012 - December 2018. Two major correlation clusters were identified, comprising several subclusters: Again WTI, Brent, and heating oil share a subcluster and appear highly correlated. However, they show no relevant relationship with the price development of natural gas futures. These results appear consistent with Figure 1. Furthermore, the crude and heating oil group shows a higher correlation with metals than with natural gas. The daily futures returns for sugar, coffee, cotton, corn, and wheat fell into the second cluster on the left, indicating the most vital relationship between corn and wheat and between sugar and coffee futures. Generally, agricultural commodities' price development does not appear to be closely related to commodities from the metals and fossil energy sector - except for natural gas.

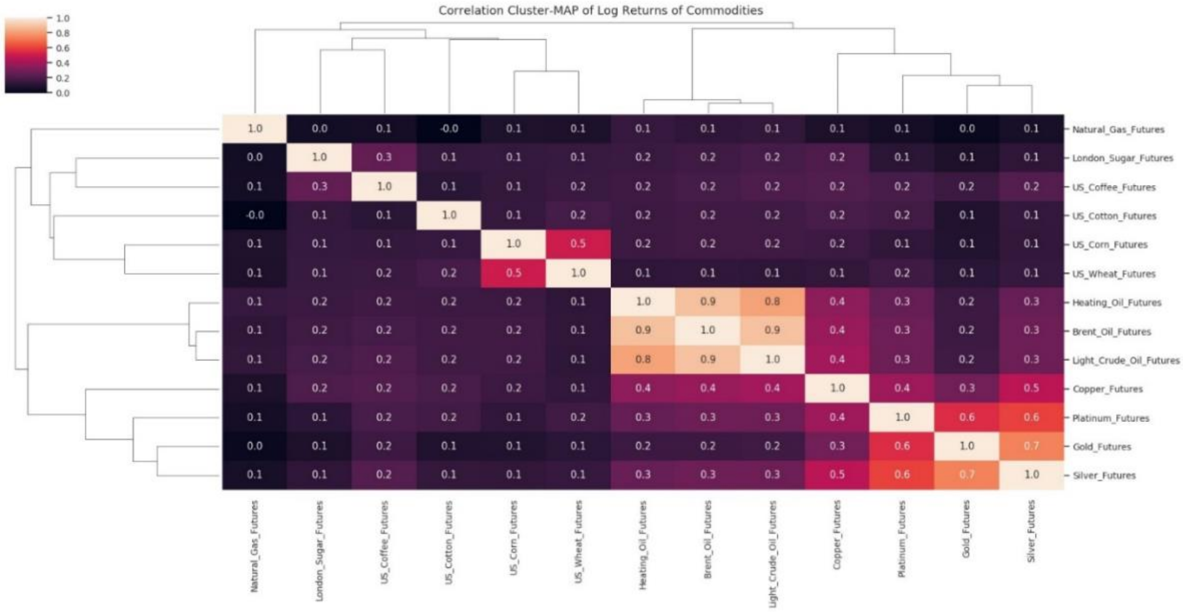

Figure 2: Correlation cluster map of daily energy, soft and hard commodity futures returns, observation period 2012-2018.

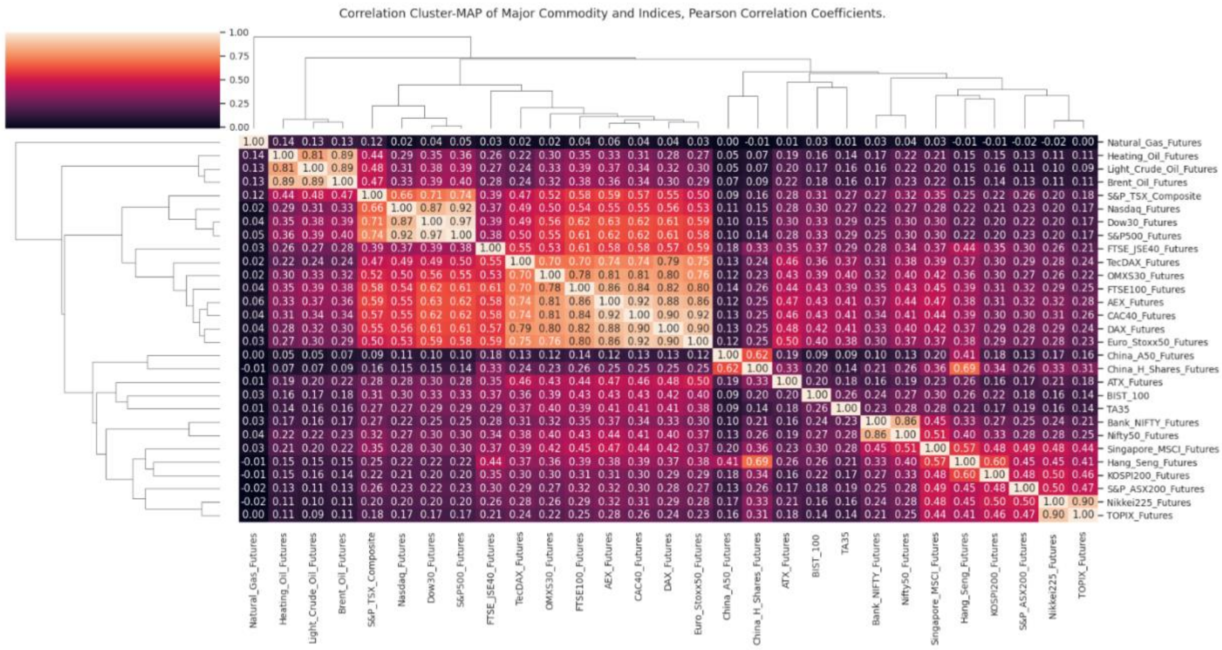

Figure 3: Correlation cluster map of global commodity futures from the fossil energy, agricultural and metals sector; daily returns over the period 2012-2018 were applied.

Global Indices and Fossil Energy Futures (Figure 3):

Figure 3 shows the correlation cluster map for a selection of 25 leading global indices and the fossil energy futures WTI, Brent, and heating oil. The daily futures returns comprised the observation period 2012-2018. The indices and abbreviations are listed in the appendix. According to the dendrogram, four relevant subclusters appear: Firstly, the South-East Asian index group including India and Australia, secondly the North-American subcluster comprising the United States and Canada, thirdly the European and Eurozone cluster including the South African Index and a final fifth cluster comprising the fossil energy futures WTI, Brent and heating oil. Again the Natural Gas Futures stand separately and are not assignable to any subcluster. Interestingly, the stock indices of Turkey (BIST100), Israel (TA35), and Austria (ATX) were grouped in one subcluster, whereby ATX shows a moderate correlation with European and Eurozone stocks.

Conclusion

In this study, the hierarchical correlation cluster mapping technique was applied to search for interrelationships and comovements among commodities (fossil energy, agricultural, and metals sectors), stocks from the fossil energy sector, and a wide range of global indices. Daily futures returns were applied over the period 2012 to 2018. The method identified clusters and correlations among the market players. The clustering outcome was visualized in the form of dendrograms and a colored correlation heat-map. Both the hierarchically branched dendrogram and the color gradients of the heat-map allow an intuitive understanding of the results and facilitate the recognition of existing correlation patterns. Additionally, the Pearson correlation coefficients were indicated on the cluster map. The results widely appear to be logically grouped; however, they revealed some novel information. Stock indices of South Korea share a subcluster with Australian stocks. Both are grouped with the Japanese Indices to form a cluster on the next higher level. However, the highest correlation coefficient is found with the Heng Seng futures. Interestingly, the South-African index shares a subcluster with the European and Eurozone indices. Among the commodities, the relatively stand-alone positioning of natural gas became apparent, which indicates an entirely surprising result. As a separate branch, it is grouped with agricultural (soft) commodities. The soft and metals commodities occupy distinct clusters as expected. Among the fossil energy stocks and commodities (Figure 1), Brent oil returns development turns out to be more closely related to WTI and US oil company stocks. The revealed interactions awaken demand for further in-depth investigations. The technique gave an outlook, which additional factors could be taken into focus. The applied method provides state-of-the-art grouping and visualization tools and an accelerated screening opportunity. Correlation clustering can be a helpful tool in the screening process for relationships of oil in the markets allowing intuitive pattern recognition. Visualization techniques are capable of offering improved reporting performance. The hierarchical clusters and correlation patterns demonstrated extended market interactions. Indications exist that the highest correlation values do not always follow the lowest grouping level. They also occur among scores separated by high hierarchical levels. Different reasons may produce a high degree of similarity in the price development, e.g., closely related economic environments, spillover effects originating from different economic branches or macroeconomic news announcements. It would be supportive to combine the method with additional techniques, e.g., machine learning and autoregressive methods like ARMAX using exogenous variables, to examine the indicated historical correlations more profoundly and to reveal real-time interactions between the market players.

Acknowledgment

For valuable support and advice, I would like to thank Professor Dr. Gerald Futschek (Vienna University of Technology), Professor Dr. Robert Kunst (University of Vienna), Dr. Hojatollah Ghanimi Fard (former Head of the Petroleum Studies Department at OPEC), Mr. Mehrzad Zamani (former Energy Models Analyst at OPEC), Professor Dr. Natalia Kryvinska (University of Vienna), Professor Dr. Michael Hauser (Vienna University of Economics and Businesses), and last but not least my patient wife Elisabeth Faseli, BSc.

Annotation

The results, tables, and figures provided in this article were part of my dissertation (Faseli, 2020, b)

References

Bakas, D., & Triantafyllou, A. (2018). The impact of uncertainty shocks on the volatility of commodity prices. *Journal of International Money and Finance*, 87, 96–111. https:/doi.org/10.1016/j.jimonfin.2018.06.001

Bakas, D., & Triantafyllou, A. (2019). Volatility forecasting in commodity markets using macro uncertainty. *Energy Economics*, 81, 79–94. https:/doi.org/10.1016/j.eneco.2019.03.016

Barberis, N., Shleifer, A., & Wurgler, J. (2005). Comovement. *Journal of Financial Economics*, 75(2), 283–317. https:/doi.org/10.1016/j.jfineco.2004.04.003

Baumohl, B. (2013). In: The Secrets of Economic Indicators, Pearson Education Inc., pp 6-12.

Becker, C., & Gather, U. (1999). The Masking Breakdown Point of Multivariate Outlier Identification Rules. *Journal of the American Statistical Association*, 94(447), 947–955. https:/doi.org/10.1080/01621459.1999.10474199

Brodinová, Š., Filzmoser, P., Ortner, T., Breiteneder, C., & Rohm, M. (2019). Robust and sparse k-means clustering for high-dimensional data. *Advances in Data Analysis and Classification*. https:/doi.org/10.1007/s11634-019-00356-9

E.I.A. (2019). What Drives Crude Oil Prices? *US Energy Information Administration*. Accessed on Dec 8, 2019. [https://www.eia.gov/finance/markets/crudeoil/index.php]

Ebrahim, Z., Inderwildi, O. R., & King, D. A. (2014). Macroeconomic impacts of oil price volatility: mitigation and resilience. *Frontiers in Energy*, 8(1), 9–24. https:/doi.org/10.1007/s11708-014-0303-0

Faseli, O., & Zamani, M. (2016). Application of Open Web API: The Impact of Crude Oil Stocks Change Announcements on Crude Oil Price Volatility, *IJEDR,* (4)1. ISSN:2321-9939.

Faseli, O. (2019). Ranking of US macroeconomic news impacting WTI crude oil volatility risk. *International Journal of Research in Business and Social Science* (2147- 4478), 8(6), 49–57. https:/doi.org/10.20525/ijrbs.v8i6.515

Faseli, O. (2020). The relationship between European Brent crude oil price development and US macroeconomy. *International Journal of Research in Business and Social Science* (2147- 4478), 9(1), 80–87. https:/doi.org/10.20525/ijrbs.v9i1.587

Filzmoser, P., Maronna, R., & Werner, M. (2008). Outlier identification in high dimensions. *Computational Statistics & Data Analysis*, 52(3), 1694–1711. https:/doi.org/10.1016/j.csda.2007.05.018

Glen, S. (2018). Ward's Method (Minimum variance method). Accessed on June 26, 2018. [https://www.statisticshowto.com/wards-method/]

Gordon, A. (1999). *Classification*, 2nd Edition. C&H/CRC Monographs on Statistics & Applied Probability. https:/doi.org/10.1201/9781584888536

Kilian, L. (2010). Oil Price Volatility. (2010). *WTO Working Papers*. doi:10.30875/9b9f6bc8-en

Kilian, L., & Vega, C. (2011). Do energy prices respond to u.s. macroeconomic news? *Review of Economics and Statistics*, 93(2), 660–671. https://doi.org/10.1162/rest_a_00086

Kilitcioglu, D. (2018). Hierarchical clustering and its applications. [https://towardsdatascience.com/hierarchical-clustering-and-its-applications-41c1ad4441a6] Accessed on Oct. 20, 2020.

Murtagh, F., & Legendre, P. (2014). Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *Journal of Classification*, 31(3), 274–295. https:/doi.org/10.1007/s00357-014-9161-z

Pierson, L. (2015). *Data Science for Dummies*. John Wiley & Sons, 81-98.

SciPyCommunity, (2020). *SciPy Reference Guide*. Release 1.5.1, 1-605.

Su, Z., Lu, M., & Yin, L. (2018). Oil prices and news-based uncertainty: Novel evidence. *Energy Economics*, 72, 331–340. https://www.doi.org/10.1016/j.eneco.2018.04.021

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58(301), 236–244. https:/doi.org/10.1080/01621459.1963.10500845

Ye, H.-M., Wang, M., & Wang, T.-L. (2011). Trading rules for high-frequency financial data based on hybrid clustering. 2011 *Eighth International Conference on Fuzzy Systems and Knowledge Discovery* (FSKD). doi:10.1109/fskd.2011.6019696

Xiao, L., & Aydemir, A. (2007). Volatility modelling and forecasting in finance. *In: Forecasting Volatility in the Financial Markets, Quantitative Finance*, 1–45. doi:10.1016/b978-075066942-9.50003-0

Zhang, X., Lai, K. K., & Wang, S.-Y. (2008). A new approach for crude oil price analysis based on Empirical Mode Decomposition. *Energy Economics*, 30(3), 905–918. https:/doi.org/10.1016/j.eneco.2007.02.012

Appendix

Figure 1 – Energy Sector: Crude Oil Stock Companies	
Anadarko Petroleum Corporation (APC)	USA
Exxon Mobil Corporation (XOM)	USA
Range Resources Corporation (RRC)	USA
Southwestern Energy Company (SWN)	USA
Valero Energy Corporation (VLO)	USA
Gazprom (GAZP)	Russia
NK Rosneft PAO (ROSN)	Russia
Lukoil (LKOH)	Russia
Oesterreichische Mineralöl Verwaltung AG (OMVV)	Austria
Total SA (TOTF)	France
Ente Nazionale Idrocarburi (ENI)	Italy
Equinor ASA ADR (EQNR)	Norway
Schlumberger NV (SLB)	Netherlands
Repsol (REP)	Spain

Figure 3 – Global Stock Market Index Futures	
Asia:	
China A50	China
China H Shares	China
Hang Seng Futures	China
Bank NIFTY	India
Nifty50	India
Nikkei225	Japan
TOPIX /Tokyo Stock Price Index	Japan
Singapore MSCI	Singapore
KOSPI200 /Korean Stock Price Ix	South Korea
TA35	Israel
BIST100 (Borsa Istanbul)	Turkey
Australia:	
S&P ASX200	Australia
Europe:	
ATX (Austrian Traded Index)	Austria
Euro Stoxx50	Eurozone
CAC40	France
FTSE100 / Financial Times Stock Exchange Index	Great Britain
DAX	Germany
TecDAX	Germany
AEX (Amsterdam Exchange Index)	Netherlands
OMXS30	Sweden
Africa:	
FTSE JSE40 (Johannesburg Stock Exchange)	South Africa
North America:	
S&P TSX Composite (Toronto Stock Exchange)	Canada
Dow30	USA
Nasdaq Futures	USA
S&P500	USA