ERASMUS UNIVERSITY ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS
International Bachelor Economics and Business Economics

Overreaction Hypothesis on crude oil prices

The formation of a profitable strategy that includes Google Trends data on dramatic events

Author: Martin Mitrov
Student number: 536258
Thesis supervisor: Dr. Jan Lemmen
Second reader: YS Gangaram-Panday
Finish date: July 2022

PREFACE AND ACKNOWLEDGEMENTS

Foremost, I want to show my gratitude and appreciation to my supervisor – dr. Jan Lemmen, who supported me during the whole process and helped me with the research on this interesting topic.

Afterwards, I thank my grandfather – Viktor Steshkin, who always believed in me and has shown me the importance of learning.

Last but not least, special thanks to prof. Christopher Baum who answered my emails and told me about the inner workings of STATA.

Working on this thesis was challenging but also extremely enjoyable. Therefore, I hope my findings will encourage future researchers.

The views stated in this thesis are those of the author and not necessarily those of the supervisor, second assessor, Erasmus School of Economics or Erasmus University Rotterdam.

ABSTRACT

This thesis investigates whether regular investors can profit from the overreaction hypothesis on crude oil prices. Google Trends data is used as a measure of investor attention, and its ability to forecast crude oil prices is assessed. The findings of the paper conclude that dramatic events are short-termed and that crude oil prices do not have a two-day rebound. Furthermore, it is deduced that crude oil prices and investor attention are negatively correlated and including the Google Search Volume Index (GSVI) does not lead to better forecasting accuracy for crude oil prices. Finally, the paper states that only including the GSVI is insufficient to create a profitable strategy for retail investors.

Keywords:
crude oil
crude oil prices
investor sentiments
Google Trends
Google Search Volume Index
overreaction hypothesis

JEL Classification:
C53
G14
G17
Q47

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Introduction

Crude oil is considered one of the most important commodities in the global market (Li et al., 2015). Its prices are influenced by a variety of factors such as aggregate supply, global oil inventory, aggregate demand and speculative demand (Gong et al., 2018). Furthermore, social media and the Internet have become an essential part of most people's lives. Individuals use the Internet for various reasons - such as information searching on Wikipedia, watching movies on Netflix or for financial advice regarding the stock market. However, in order to search for the previously mentioned queries, individuals use search engines, and the most used one is called Google. This engine has an extension called Google Trends, and the data collected from it is called Google Search Volume Index (SVI). This extension shows data of various search queries and their popularity during different time periods. Li et al. (2015) state that the SVI reflects the emotions of retail investors who overreact in times of crisis and are more irrational and emotional. Therefore, my study will examine whether the behaviour of retail investors can create abnormal returns in crude oil prices. The thesis will review recent major events that have caused turmoil and will conclude whether those events have caused regular investors' overreaction. Hence, in this paper, I engage in a discussion regarding the overreaction hypothesis on crude oil prices and have formulated the following research question:
"Can retail investors profit from the overreaction hypothesis on crude oil prices?"

The foundation of the overreaction hypothesis originates from applied psychology, and it states that people tend to overreact to dramatic news, regardless of whether they are positive or negative (Mun et al., 2000). The topic of investor overreaction is first identified by John Maynard Keynes in his book "The General Theory of Employment, Interest and Money" (1936/2018, p. 135). He says that it is absurd how day-to-day fluctuations are excessive and that during holidays the market valuation of the British railway system spikes. Furthermore, de Bondt and Thaler (1985) were one of the first to conduct a study on market overreaction; however, their study is about stocks and not oil prices. They write that stocks which have exhibited positive abnormal returns are followed by a few consecutive days of negative returns, while stocks that have exhibited negative abnormal returns are followed by positive ones. Bremer and Sweeney (1991) build up on the idea and state that an event occurs when there is a daily return that is lower than -10%. However, a study done by Cox and Peterson (1994) says that the bid-ask bounce causes most of the return reversal and investor reaction has a way smaller effect on it. This paper concludes that the market overreaction is not as effective as previous papers have stated and that it is hard to profit from it. Lastly, two papers written by Guo and Ji (2013) and Li et al. (2015) research the data of Google Trends and if it can be used as a predictive tool for long-term oil prices.

The implications of this paper can be used by retail investors to understand whether crude oil prices are affected by overreaction or mostly by supply shortages caused by periods of financial turmoil, wars, or natural disasters. If the paper proves that market overreaction affects crude oil prices, then regular investors can predict the potential longevity of a crisis by using Google Trends and to possibly generate profits. On the other hand, if overreaction is proven to have no significant effect on crude oil prices, then investors will know that factors such as investor sentiments and datasets like Google Trends are not that useful for forecasting crude oil prices.

Moreover, this thesis can be used for academic purposes since no paper of the aforementioned ones has clearly answered my research question. The other papers talk about overreaction on stocks and not on commodities; therefore, it is unclear if overreaction can cause crude oil prices to change and if retail investors can profit from it. Also, this paper can be used by other researchers to compare their findings with mine and to further investigate the overreaction hypothesis and its influence on commodities such as crude oil.

This work combines the topics of crude oil prices, the overreaction hypothesis and the Internet, which has never been done before. Most of the papers have either researched the overreaction on stocks or on other securities different from crude oil. Other studies like Guo and Ji (2013) and Li et al. (2015) have researched whether the Internet and more particularly Google Trends can predict crude oil prices. However, they do not exclusively investigate for events that have caused a daily drop in crude oil prices by -10% or a daily increase of 10%. My paper combines all the mentioned factors and research whether investors can profit from a daily price change of 10% or more. Also, it finds if there are follow-up positive price changes for negative abnormal returns and negative price changes for positive abnormal returns, as well as if those changes are enhanced by investor reactions.

This paper finds that there are 37 overreactions for WTI prices of which 14 are negative, and 23 are positive. Furthermore, for the studied period, Brent crude oil prices have 26 overreactions, the negative and positive being 11 and 15, respectively. It is also concluded that opposite to stocks, crude oil prices do not have a two-day rebound. Moreover, that dramatic events are short-termed, and the SVI significantly decreases a year after the end of the event. Finally, this thesis finds a negative correlation between SVI and crude oil prices, as well as that only the competing model for Brent crude oil prices leads to slightly higher forecasting accuracy.

The remainder of this study is organised in the following order. Section 2 discusses the literature on the overreaction hypothesis, the Internet and crude oil prices. Section 3 describes the data used in this paper. Section 4 shows the methods implied. Section 5 depicts the results that I have found from doing my research, and section 6 concludes the thesis.

Chapter 2 Literature Review

This chapter examines the existing literature, provides definitions and discusses the results of previous studies. Furthermore, it assesses papers related to the overreaction hypothesis, investor attention and oil prices.

Furthermore, Table 2.1 is a meta table that outlines all research papers that this thesis discusses. Also, it briefly summarises the time period, region, methods used, if there are any control variables and the results of those studies. The meta table provides with a timeline that sheds light on the results and the degree to which they have changed during the years. From analysing the table, it can be concluded that articles that consider the SVI and crude oil prices are relatively new. This is in line with the fact that Google Trends only captures data newer than 2004. Furthermore, it shows us that when authors research the overreaction hypothesis on stock prices, their control variable is "Firm size", the only exception being Cox and Peterson (1994). Oppositely, the papers that study investor attention and its influence on crude oil prices, do not use any control variables apart from Loughran et al. (2019) who have taken a different approach.

Table 2.1

Meta table that describes the literature used for the thesis

Author(s) (Publication year)	Time period	Region	Method	Control variables	Results
De Bondt and Thaler (1985)	January 1926 – December 1982	US	Event study market model	Firm size	$[ACAR_{L,36} - ACAR_{W,36}] = 0.246$
Bremer and Sweeney (1991)	1962-1986	US	Event study	Firm size	Average $rebound_{day\ 1} = 1.773\%$ Cumulative $rebound_{day\ 2} = 2.2\%$
Cox and Peterson (1994)	January 1963 - June 1991 (excluding	US	Event study	Bid-ask bounce	No evidence supporting the overreaction hypothesis.

	September & October 1987)				
Clare and Thomas (1995)	January 1955 - 1990	UK	Event study	Firm size and seasonality	Losers outperform previous winners for two years with 1.7% annually. Losers tend to be small firms.
Ellen and Zwinkels (2010)	January 1983 – August 2009	Worldwide	Heterogenous agents model, VAR model and random walk.	No control variable	Speculators often switch between the fundamentalist and chartist models.
Vansteenkiste (2011)	January 1992 – April 2011	Worldwide	Markov switching models with time-varying transition probabilities	No control variable	Large unexpected shocks in crude oil prices lead to all traders entering the market.
Guo and Ji (2013)	2nd of January 2008 – 18th of November 2011	Worldwide	Event study, Granger causality test	No control variable	The Internet influences short-term oil prices. An existing long-term relationship between Brent prices and market concern.
Li et al. (2015)	January 2004 – June 2014	Worldwide	Event study, Granger causality test	No control variable	Google Trends improves the forecast accuracy for a week ahead and measures investors' attention for non-commercial traders.
Han et al. (2017)	January 2004 –	Worldwide	EEMD hybrid approach and WN model	No control variable	Investor attention can forecast daily

	March 2016				and weekly crude oil prices.
Yao et al. (2017)	January 2004 – November 2016	Worldwide	Principal component analysis approach, SVAR model.	No control variable	Investors can profit from crude oil price fluctuations. Investor attention contributes to 15.18% of WTI crude oil price fluctuations.
Yao and Zhang (2017)	January 2004 – June 2016	Worldwide	Augmented DF, Phillips, and Peron, ARIMA model.	No control variable	Google index negatively impacts crude oil prices and cannot forecast them.
Loughran et al. (2019)	January 2010 – September 2016	US	OLS regression and GARCH models used.	Gold spot price, trade-weighted US dollar index, number of daily oil articles and etc.	Media affects security prices. Higher number of negative words leads to lower short-term prices.
Borgards et al. (2021)	20th of November 2019 – 3rd of June 2020	Worldwide	Event study, dynamic model approach, log prices used.	No control variable	Crude oil prices, differently from other commodities, has more negative overreactions than positive ones during the Covid-19 pandemic.
Haque and Shaik (2021)	20th of February 2020 – 27th of April 2020	Worldwide	GARCH, ARMA and ARIMA	No control variable	ARIMA (4,1,4) is the best model, especially in times of crisis.

2.1 Overreaction hypothesis

As mentioned in the introduction, de Bondt and Thaler (1985) are one of the first to assess the overreaction hypothesis on the market. Their study focuses on stock prices and if they are affected by the overreaction of investors caused by dramatic events. The paper concludes that portfolios of past losers have subsequently earned 24.6% more in the future than those of past winners. Also, it indicates a possible market overreaction since it mentions abnormal returns that occur every January for 5 consecutive years. Similar results were achieved by Clare and Thomas (1995) that conclude that losers outperform winners over a two-year period by approximately 1.7% annually. Furthermore, they claim that the reason for the overreaction effects might be due to firm sizes, which was not previously deduced by de Bondt and Thaler (1985). Another study by Bremer and Sweeney (1991) says that a dramatic event occurs when prices decline or increase by 10% or more. Additionally, they conclude that for stocks after a large negative daily drop caused by a dramatic event, there is a follow-up of two days of positive returns. The authors say that for a -10% trigger, the average day 1 rebound is 1.773% and cumulative day 2 is 2.2%. Also, they infer that market prices usually do not adjust as quickly as expected, which hints that the overreaction hypothesis might not exist. Cox and Peterson (1994) support this claim by concluding that there is a significant rebound after a large drop that is rather caused by the bid-ask bounce and not by investor sentiments. Therefore, after examining the last papers, my first hypothesis is:

"*Dramatic events cause a rebound in crude oil prices.*"

Borgards et al. (2021) examine the overreaction behaviour of 20 commodities during the Covid-19 pandemic. The difference is that this paper does an intraday analysis instead of a daily, weekly, or yearly one. The data is collected from 20th of November 2019 to 3rd of June 2020, which is useful since I can compare my findings to the ones of this study. The authors believe that such an analysis would allow investors to make quicker decisions since the market changes rapidly. One of the commodities used in Borgards et al. (2021) is crude oil, and it is concluded that this commodity has a different overreaction behaviour from other commodities. The study concludes that crude oil has had more negative overreactions than positive ones during the Covid-19 pandemic. Furthermore, it believes that this is useful information for investors, which should probably expect similar behaviour in future lockdowns or dramatic events. Therefore, my second hypothesis is:

"Negative overreactions were more than the positive ones during the Covid-19 pandemic, and there was a similar pattern during other times of crisis."

2.2 Google Trends & The Internet

Google Trends has been used by several papers to evaluate market sentiments and its correlation to crude oil prices. Guo and Ji (2013) analyse both short-run and long-run market concerns by using SVI. They conclude that dramatic events such as the Libyan War have a major asymmetric effect on short-term crude oil price volatility. This effect is usually strengthened by the Internet and affects all types of crude oils. However, it might affect certain types of oil more than others. For instance, during the Libyan War, Brent crude oil prices were affected more than WTI ones. The reason for this is that Libya was mainly exporting to Europe which led to investor concern in this local market that quickly triggered an increase in European crude oil prices. Furthermore, the study deduces that those short-run market concerns only exist in specified periods, which is shown in figure 2.1, and they quickly diminish after the crisis ends. Lastly, it is concluded that the Internet has begun playing a big role in influencing crude oil prices and that it is representative of current market concerns. Thus, my third hypothesis follows after assessing the findings of Guo and Ji (2013): *"Dramatic events have a short time effect on investor attention."*

Li et al. (2015) is another paper that uses the SVI to measure its relation to crude oil prices. However, this paper does not concentrate on crude oil price volatility as much as on Google Trends, as a tool that measures investor attention, different trader positions and crude oil forecast accuracy. This work concludes that the SVI captures the investor attention of retail traders but not of commercial ones. Furthermore, it is mentioned that investor attention usually drives price volatility, and that price volatility precedes investor attention for two weeks. The aforementioned implies that investor attention is correlated with price volatility, and usually one affects the other. Lastly, the paper concludes that the search volume index can be used to create forecasts and that it is more precise than many other models.

Furthermore, Yao et al. (2017) investigate the influence of investor attention on crude oil prices. In their analysis, they use 11 keywords, some of which are "oil price" and "current oil price". Moreover, the study creates a four-variable SVAR model that estimates the degree to which WTI crude oil prices respond to investor attention. Similar to Li et al. (2015), this study suggests that Google Trends reflects the choices of retail investors which are regarded as more emotional and irrational. Also, Yao et al. (2017) say that commercial investors are less dramatic and more attention-driven. The paper believes that exactly these investors wait for the dramatic event to end and then start putting pressure on crude oil prices in order to earn profits. For example, after the financial crisis in 2008, the crude oil prices decreased, but they started to increase again in February 2009 which was largely caused by institutional investors. Finally, the conclusion of the study is that around 15.18% of WTI crude oil price fluctuations are triggered by investor attention.

Additionally, Friedman (1953) argues that rational investors trade against irrational ones by taking the opposite position, which usually drives prices back to fundamentals. This is in line with the findings of Yao et al. (2017) and the conclusion that exactly rational traders put pressure on crude oil prices at the end of a dramatic event which leads to a price increase. The previous two sentences spark a discussion about fundamentalists versus chartists and how the different types of trader groups affect crude oil prices. According to Ellen and Zwinkels (2010), fundamentalists are traders who believe that market prices return to their fundamental value, and the difference between the actual price and the fundamental value is a profit opportunity. Oppositely, chartists base their expectations on past prices and anticipate trends to continue in the same direction, which drives prices away from the fundamental value and destabilises the market. Furthermore, the paper claims that both type of groups influence crude oil prices; however, often speculators switch between them based on past profitability. Vansteenkiste (2011) supports the idea and says that up to year 2004, movements in oil prices were best described by underlying fundamentals, while after 2004 switching between the two groups has become more frequent. This made the market more unstable since more investors were following the chartist model and crude oil prices started to significantly fluctuate from their fundamental values. Moreover, Yegorov (2009) makes a similar deduction and states that the growth of chartists presence is correlated with the growth of market instability. So far, no paper has examined whether Google Trends can indicate the periods where one of the strategies is more prevalent than the other. However, as Loughran et al. (2019) have concluded, retail investors can easily be influenced by news articles. This implies that by following the news, retail investors are the cause of market overreactions since they enhance the effect of the strategies followed by institutional investors. Vansteenkiste (2011) reaches a similar conclusion by stating that when the chartist strategy prevails over the fundamentalist one, then unexpected shocks in crude oil prices occur, mainly due to all traders entering the market. Lastly, the discussion of fundamentalists versus chartists is important since it helps to better understand why the market has destabilised and overreactions are caused.

Furthermore, Loughran et al. (2019) take a different approach from the other papers. It uses a set of keywords divided into three different categories: the first two are terms that are associated with an increase and decrease in oil prices, while the third one is a category with terms that depend on the modifier. Those keywords are collected from various news sources, and the idea is to check whether the emotion that a certain source implies can have an impact on its readers. The authors conclude that higher counts of negative keywords in an article lead to lower crude oil prices on the same day. This shows that readers which are mostly retail investors are often irrational and tend to overreact to news. Even though Loughran et al. (2019) research the emotions that different news sources imply and their effect on crude oil prices, it proves the theory of Li et al. (2015) and Yao et al. (2017) that retail investors are irrational. The article is of use for my thesis since it strengthens my notion that Google Trends represents the current emotions of retail investors. Furthermore, it helps in answering the question of whether retail investors can profit from monitoring the SVI. Finally, the rise of the chartist model, Loughran et al. (2019), Yao et al. (2017) and their study of investor attention, as well as Li et al. (2015) and their correlation analysis complement each other and lead to my fourth hypothesis:
"A crisis can be identified by monitoring the Google Trends volume."

It is largely known from previous studies that Google Trends primarily represents the sentiments of retail investors, who are considered more irrational and emotional. Therefore, by answering this hypothesis, it will also be concluded whether only using the SVI is a profitable strategy. Thus, the fourth hypothesis helps formulate the fifth one which is:
"If a crisis is identified, it is a profitable strategy to buy crude oil whenever the SVI volume starts to decrease."

An academic article by Han et al. (2017) evaluates the forecasting power of crude oil prices by using the SVI. It collects 1124 keywords which it divides into three groups. The first group of attention terms is based on variables that describe oil, such as "crude oil", "crude oil price" and "WTI". The second and third groups reflect terms that measure investor attention in financial markets and in fundamentals, respectively. The authors claim that this is necessary since changes in financial asset returns are correlated with crude oil price movements. They conclude that by using their keywords, precise forecasts for oil price movements can be created. Furthermore, that their models – the Westerlund and Narayan (WN), and the ensemble empirical mode decomposition (EEMD) hybrid approach perform well in predicting

oil prices when investor attention is used. Another paper by Yao and Zhang (2017) examines whether the SVI can help improve the forecasting performance of WTI crude oil prices. It uses four models – ARIMA, ARMAX, ARMA-GARCH, ARMAX-GARCH, and gets opposite results to those of Han et al. (2017). Yao and Zhang (2017) conclude that the Google Index has a negative impact on crude oil prices, but it does not improve the price forecast performance. This paper, as well as the ones by Han et al. (2017) and Li et al. (2015), lead to my last hypothesis, which is about forecasting:
"The forecasting accuracy of crude oil prices improves when the SVI is included."

Additionally, to support my research, I use a paper written by Haque and Shaik (2021). This study searches for the model that is the most accurate in predicting crude oil prices in times of turmoil. It goes through various types of GARCH models, ARMA models and the ARIMA model. Haque and Shaik (2021) conclude that the ARIMA (4,1,4) model is the best for predicting crude oil prices. Furthermore, it claims that ARIMA which is the extended model of ARMA is better since it removes non-stationarity. Lastly, the paper mentions that series are stationary when two consecutive values depend only on the time interval between them and not on the time itself.

Chapter 3 Data

The data for the crude oil prices for both Brent and WTI is collected from Bloomberg Terminal. Furthermore, in order to be sure that my data is precise, I compared it to the data from the U.S. Energy Information Administration (EIA) which is used by He et al. (2021) for WTI prices as well as by Li et al. (2015) for Brent prices.

3.1 Oil prices data

Table 3.1 illustrates the descriptive statistics of crude oil average daily returns. It consists of three columns, two representing the WTI average daily returns and one representing the Brent average daily returns. I have created two columns for the WTI average daily returns because of the min and max observations which are -306% and 127% respectively. This, together with the very high skewness and kurtosis implies that there might be an outlier that significantly affects the descriptive statistics. This outlier is an event that was observed on 20/04/2020 when WTI prices dropped to -$37.63. Therefore, to have a better comparison between WTI and Brent's average daily returns, I have decided to remove the 20/04/2020 observation from the dataset.

After removing the outlier, the skewness and kurtosis of the WTI become similar to this of the Brent average daily returns. Furthermore, the mean for WTI in the period 06/01/2004-20/04/2022 is 0.0635%, while for Brent, it is 0.0535%. Lastly, the median implies that the average daily returns of WTI are unaffected by the removal of the outlier and are significantly higher than those of Brent, being 0.1079% and 0.0525%, respectively.

Table 3.1

Descriptive statistics of the crude oil average daily returns (06/01/2004 - 20/04/2022)

	WTI average daily returns	WTI average daily returns (without 20/04/2020)	Brent average daily returns
Mean	0.0342%	0.0635%	0.0535%
Median	0.1084%	0.1079%	0.0525%
Std. Deviation	0.0559	0.0278	0.02346
Min	-305.9660%	-45.2107%	-28.5304%
Max	126.6011%	28.2573%	21.8689%
Skewness	-33.2659	-0.1421	-0.2170

Kurtosis	2024.8373	29.6676	13.1388

3.2 Google Trends data

Furthermore, for the Google Trends data, four keywords are used, namely: “crude oil”, “crude oil price”, “WTI oil”, and “Brent oil”. When choosing the keywords, I had the option to use the search words as terms or topics. I chose to use them as search terms because for topics the results are broader, and the information might overlap with the other keywords. Precisely, topics are a group of terms that share a similar concept, while search terms are individual terms that include only queries that have had the keywords in them.

Table 3.2 shows the monthly descriptive statistics of the keywords used in Google Trends from January 2004 to April 2022. For the WTI oil and the Brent oil queries, some of the data were depicted as “<1” which is considered non-numeric by Excel. Therefore, to perform the descriptive statistics, I have replaced the data shown as “<1” with “0.5”. I chose to replace it with “0.5” because it is the average of 0 and 1, and since it is unknown whether the number is closer to 0 or to 1, it is best to pick the average of the two. Additionally, the table shows that the search term “crude oil” has the highest mean of all, being 18.34. This means that this keyword is the most searched of the four on Google. Moreover, Table 3.3 displays the correlations between keywords used on Google. The correlations between “crude oil” and the other three keywords are 0.9484, 0.773 and 0.8552 for “crude oil price”, “WTI oil” and “Brent oil” respectively.

Table 3.2

Monthly descriptive statistics of the keywords used in Google Trends (January 2004 - May 2022)

	Crude oil	Crude oil price	WTI oil	Brent oil
Mean	18.3484	5.9593	1.2557	2.2738
Standard Error	0.7099	0.7683	0.5300	0.7975
Median	16	5	1	2
Std. Deviation	10.5534	5.7091	1.8529	2.1375
Min	7	1	0.5	0.5
Max	100	50	24	18
Skewness	3.2615	3.7728	8.8721	4.1334

Kurtosis	18.0142	21.0908	103.9270	23.7314

Table 3.3

Correlation matrix of the keywords used in Google Trends (monthly data)

	Crude oil	Crude oil price	WTI oil	Brent oil
Crude oil	1	0.9484	0.773	0.8552
Crude oil price	0.9484	1	0.8178	0.9237
WTI oil	0.773	0.8178	1	0.8398
Brent oil	0.8552	0.9237	0.8398	1

Table 3.4 illustrates the daily descriptive statistics of the keywords used in Google Trends from 6th of January 2004 to 20th of April 2022. Same as for the monthly data, the daily data table indicates that the mean for the term "crude oil" is the most searched one on Google, being 11.066. Moreover, Table 3.5 depicts the daily correlations between "crude oil" and the other three terms, which are 0.8170, 0.5450 and 0.4721 for "crude oil price", "WTI oil" and "Brent oil" respectively. The significant drop in the correlations for "WTI oil" and "Brent oil" is caused by the fact that many observations are shown as "<1". Oppositely, for the terms "crude oil" and "crude oil price", the correlation does not drop significantly since there are no non-numeric observations.

Therefore, from the monthly dataset and the correlation between "crude oil" and "crude oil price" in the daily dataset, it can be concluded that the keywords are correlated and depict similar trends. Thus, when comparing the crude oil prices to the Google Trends data, only the data for the term "crude oil" can be used.

Table 3.4

Daily Descriptive statistics of the keywords used in Google Trends (05th of January 2004 – 20th of April 2022)

	Crude oil	Crude oil price	WTI oil	Brent oil
Mean	11.0660	1.3292	0.0635	0.2078
Standard Error	0.0808	0.0203	0.002	0.0049
Median	9.6	0.75	0.02	0.09
Std. Deviation	6.6057	1.6588	0.1513	0.3997

Min	0	0	0	0
Max	100	34	7.68	16
Skewness	2.1870	4.4351	22.9534	20.3808
Kurtosis	11.9115	46.4513	1022.9094	694.1833

Table 3.5

Correlation matrix of the keywords used in Google Trends (daily data)

	Crude oil	Crude oil price	WTI oil	Brent oil
Crude oil	1	0.8170	0.5450	0.4721
Crude oil price	0.8170	1	0.6276	0.5803
WTI oil	0.5450	0.6276	1	0.4376
Brent oil	0.4721	0.5803	0.4376	1

3.3 Google Trends limitations

The main limitation of the SVI is that it provides daily search frequency data only for the last 9 months. If the timeframe is larger than 9 months but smaller than 5 years, then the weekly frequency data is provided. Lastly, if the period is larger than 5 years then SVI gives us only monthly data.

To get daily data for all months, I downloaded separately the data for each month from January 2004 to April 2022. However, this data tracks only the search volume for the particular month which is independent from the volumes of the previous months. Therefore, to relate the months to each other, I multiplied the value of each day of a single month by the value of the same month from the monthly data. For instance, the value of the "crude oil" keyword for January 2004 in the monthly data is 14. Thus, I multiplied the daily values for 1^{st} of December 2004, 2^{nd} of December 2004, 3^{rd} of December 2004 up to 31^{st} of December 2004 by 0.14.

Another limitation of the data is that the "WTI oil" and "Brent oil" keywords have many observations equal to 0 in the period 2004-2014. This implies that they were not as searched as the other two keywords in the given timeframe. Furthermore, Table 3.6 depicts how the correlations have changed in the recent years implying that due to the growing interest in oil, the search volume has increased. The table infers that the correlation in the period 6th of January 2004 – 1st of January 2014 between “crude oil” and “WTI oil” is only 0.0436. Oppositely, the correlation between the two soars to 0.7405 in the period 1st of January 2020 – 20th of April 2022. Therefore, it can be concluded that the search volume in the first 10 years of the dataset for the keywords “WTI oil” and “Brent oil” is too low to imply that there is no correlation between those keywords and “crude oil”. Also, the data for the other periods indicates that whenever the search volume increases, the correlation does as well, which shows that the data for “crude oil” can be used to represent all the keywords.

Table 3.6

Correlation of “crude oil” with WTI oil and Brent oil

Period	WTI oil	Brent oil
6th of January 2004- 20th of April 2022	0.5456	0.4730
6th of January 2004 – 1st of January 2014	0.0436	0.3430
1st of January 2014 – 20th of April 2022	0.6586	0.5185
1st of January 2020 – 20th of April 2022	0.7405	0.8456

Chapter 4 Method

The formula used for average daily returns is the following:

$$Average\ Returns_t = (\frac{Price\ t}{Price\ t-1}) - 1$$

However, this formula is not feasible for negative prices since it gives negative returns for 21/04/2020. This happens because the price on 21/04/2020 is positive, while on 20/04/2020 it is negative, therefore when dividing a positive by a negative number, we still get a negative average return. To adjust this, I have put the fraction in a module, and have gotten the following equation: $AR = \left|\frac{Price\ t}{Price\ t-1}\right| - 1$. This formula was written in Excel with an IF function. Also, it is used for all average daily returns and has restrictions that automatically check for the sign of the price and whether the numerator is larger than the denominator or vice versa. I decided to use an IF function rather than a module one because a module function always makes the fraction positive, and this is not the case since the average daily returns for 20/04/2020 need to be negative, because the price for the trading day before this event was positive and smaller than the absolute price of 20/04/2020. Thus, by using a module and not the IF function, I would have gotten positive instead of negative returns. Finally, there is more information in the appendix on the outcome if the average daily return formula is used instead of an IF function.

My first hypothesis is: "*Dramatic events cause a rebound in crude oil prices."*
To answer it, I first find the events that have exhibited more than 10% of positive or negative average daily returns. In my thesis, they are found by a binary Excel function that returns "1" for days that have abnormal returns and "0" for days that do not have ones. The sum of the numbers depicted as "1" in this column is the total number of overreactions. Similar to Bremer and Sweeney (1991), only a 2-day rebound is checked. Therefore, to examine if there is a rebound, I use the "Advanced Filter" option in Excel because it filters multiple columns simultaneously and sets a list of criteria for them.[1]

[1] More information in the appendix

The second hypothesis, which derives from the findings of Borgards et al. (2021) states the following: *"Negative overreactions were more than the positive ones during the Covid-19 pandemic, and there was a similar pattern during other times of crisis."*
First, for this hypothesis I use the period from 20th of November 2019 to 3rd of June 2020. This comes from the fact that the study of Borgards et al. (2021) uses the aforementioned time frame. To find whether there are more negative than positive overreactions, I use the data for the first hypothesis. Furthermore, I investigate the Global Financial Crisis which lasts from 15th of September 2008 to 26th of June 2009 and the Libyan War from 22nd of February 2011 to 1st of November 2011 (Guo & Ji, 2013).

The third hypothesis comes from Guo and Ji, 2013 and is: *"Dramatic events have a short time effect on investor attention."*
To answer this hypothesis, I examine the Google Trends data. Moreover, the same beginning dates as for hypothesis 2 are used, while the ending period is exactly 1 year. Namely, from 20th of November 2019 to 19th of November 2020 for the Covid-19 pandemic, from 22nd of February 2011 to 22nd of February 2012 for the Libyan War and 15th of September 2008 to 15th of September 2009 for the Financial Crisis. Also, to check whether the effect is only for the short term, I examine the data for the time periods with the same length as the corresponding dramatic events. Finally, the short-term effect implies that during those time periods, the SVI index should be lower since the events are over.

My fourth hypothesis is based on the papers of Li et al. (2015) and Yao et al. (2017) and is: *"A crisis can be identified by monitoring the Google Trends volume."*
It is useful to find the correlation between crude oil prices and the SVI in order to conclude whether a retail investor can identify a crisis by monitoring the Google Trends Volume. Finding the correlation helps examine whether increased concern leads to higher or lower prices. However, to find it, it is necessary to use an Excel function once again. This happens because some crude oil days are without a given price and the SVI needs to have the same number of observations as crude oil prices. Therefore, to match the data I use an "INDEX" function and inside of it a "MATCH" function with a lookup value equal to the date of the WTI price, and a lookup array equal to the date of the keyword "crude oil". The same process is executed for WTI and Brent prices in order to check for significant differences in the correlations. Moreover, the conclusion of Loughran et al. (2019) also relates to my hypothesis. They infer that an increase in the SVI assumes negative market concerns. To

check that, I conduct an ordinary least squares (OLS) regression that finds the effect that the independent variable – Google search volume index (SVI) has on Crude oil prices of both WTI and Brent. The two regression equations are the following:

1. $WTI\ Prices_t = a_t + a_2\ SVI_t + \varepsilon_t$
2. $BRENT\ Prices_n = a_n + a_2\ SVI_n + \varepsilon_n$

For both regressions, ε_t and ε_n are the error terms respectively for WTI and Brent prices, while a_t and a_n are the constants. Lastly, the OLS analysis is conducted by assuming that the necessary assumptions are satisfied.

The follow-up hypothesis to the previous one is: *"If a crisis is identified, it is a profitable strategy to buy crude oil whenever the SVI volume starts to decrease."*
To answer this hypothesis, first I need to have an explicit answer to the previous one. If the results show that the correlation between SVI and crude oil prices is negative, then the trader can expect that a decreasing SVI would lead to higher prices and therefore he can buy. Oppositely, if the correlation is positive, then a decreasing SVI would lead to decreasing prices and it is better to short the position.

My last hypothesis has societal relevance since it allows investors to follow my steps and forecast future crude oil returns. It says: *"The forecasting accuracy of crude oil prices improves when the SVI is included."*
First, in order to know whether the SVI is useful for forecasting crude oil prices, I have constructed two models based on the steps that Tse (1997) uses in his analysis. For both, ARIMA (p,d,q) model is applied and more particularly the ARIMA (4,1,4) since it is considered the best in predicting future values especially in times of crisis (Haque and Shaik, 2021). In the model, p is the number of autoregressive terms, d is the degree of differencing, and q is the order of the moving average (MA) model. Furthermore, the authors of the paper believe that crude oil prices data is non-stationary which is the reason for ARIMA to be chosen over ARMA. The first model is the benchmark one, which does not use the SVI and only checks for the predictability of crude oil prices by using its past values. The competing model is an ARIMAX (4,1,4), where the "X" stands for "with exogenous variable" that in my analysis is the SVI. Lastly, to conduct my study, I use STATA and choose the best model by finding the one with the lowest root mean squared error (RMSE), mean absolute error (MAE) and the mean absolute percentage error (MAPE). I evaluate the results by using a Diebold-Mariano (DM) test that checks whether the forecast accuracy of the competing model is

significantly different from the one of the benchmark model. Furthermore, since it is complicated to find the above-mentioned values in STATA, I work with the “fcstats” and “dmariano” add-ins that calculate the tests automatically (Baum, 2018).

Chapter 5 Results

The total number of overreactions for WTI and Brent crude oil spot prices are 37 and 26, respectively. For WTI, 14 are negative overreactions and 23 are positive ones, while for Brent crude oil prices 15 are positive and 11 are negative.

My first hypothesis is the following:

"*Dramatic events cause a rebound in crude oil prices.*"

Table 5.1 shows that for WTI crude oil prices, only 11 overreactions have caused a reverse 2-day rebound. Moreover, one rebound is positive after a negative overreaction and 10 are negative after a positive overreaction. Therefore, I reject the hypothesis for a reverse rebound after WTI crude oil price overreactions. However, I encourage further research with a larger data sample since 43.5% of the positive overreactions have caused a reverse rebound. Furthermore, for Brent crude oil prices, there are four positive rebounds and four negative ones after negative and positive overreactions, respectively. Therefore, since the rebounds are too few, I also reject the hypothesis for Brent crude oil price overreactions.

Table 5.1

Number of rebounds for negative and positive overreactions

	Total overreactions	Rebound after a negative overreaction	Rebound after a positive overreaction
WTI Crude oil prices	37	1	10
Brent Crude oil prices	26	4	4

My second hypothesis states that:

"Negative overreactions were more than the positive ones during the Covid-19 pandemic, and there was a similar pattern during other times of crisis."

Table 5.2 depicts the total number of positive and negative overreactions during major dramatic events, such as the Covid-19 pandemic, the Libyan War, and the Global Financial Crisis. Those events have caused more overreactions than all the other periods combined. Overall, 30 out of 37 and 20 out of 26 overreactions are generated during the aforementioned events for WTI and Brent prices, respectively. Furthermore, Guo and Ji (2013) state in their analysis that the Libyan War has caused major volatility fluctuations, especially in the European crude oil market. However, according to my analysis, those fluctuations were not enough to create a single market overreaction. Lastly, I reject my hypothesis since the positive overreactions are more than the negative ones.

Table 5.2

The number of negative and positive overreactions during major dramatic events

	WTI Crude oil prices		Brent Crude oil prices	
Dramatic Event (Period)	Negative Overreactions	Positive Overreactions	Negative Overreactions	Positive Overreactions
Covid-19 Pandemic (20/11/2019 – 03/06/ 2020)	7	11	7	7
Libyan War (22/02/2011-01/11/2011)	0	0	0	0
Global Financial Crisis (15/09/2008-26/07/2009)	4	8	2	4

The third hypothesis in my thesis is:

"Dramatic events have a short time effect on investor attention."

Table 5.3 depicts the averages of the keyword 'crude oil' during major dramatic events. My data covers the period from 06/01/2004 to 20/04/2022, and the average of the keyword is 11.06. Furthermore, during the Covid-19 pandemic, the Libyan War and the Global Financial Crisis, the SVI is 12.30, 7.51 and 15.72, respectively. The averages for the post-Covid, post-Libyan War and post-Financial Crisis periods are respectively 10.71, 6.74 and 8.23. My findings show that dramatic events do have a short-term effect on investor attention, and this effect declines a year after the event. All the dramatic events have higher SVIs than the time intervals representing the post periods. Furthermore, the highest decrease was in the Global Financial Average variable, which dropped from 15.72 to 8.23. Therefore, I do not reject my hypothesis and conclude that dramatic events do have a short-term effect on investor attention.

Table 5.3

Averages of the keyword "crude oil" during major dramatic events

	Total Average (06/01/2004 – 20/04/2022)	Covid-19 Pandemic (20/11/2019 – 19/11/2020)	Post Covid-19 (20/11/2020 – 20/11/2021)	Libyan War (22/02/2011 - 22/02/2012)	Post-Libyan War (23/02/2012 – 22/02/2013)	Global Financial Crisis (15/09/2008 - 15/09/2009)	Post-Global Financial Crisis (16/09/2009 – 16/09/2010)
Crude oil	11.06	12.30	10.71	7.51	6.74	15.72	8.23

My fourth hypothesis states the following:

"A crisis can be identified by monitoring the Google Trends volume."

Table 5.4. shows that by keeping everything constant, a unit increase in the SVI leads to a decrease of approximately 1.1023 in WTI prices. Furthermore, keeping all factors constant, a unit increase in the SVI would lead to a decrease of around 1.3116 in Brent prices. The following conclusion shows that the SVI and crude oil prices are related and that in times of turmoil, crude oil prices are expected to fall since the SVI increases. Additionally, the negative correlation is proven in Table 5.6. in the appendix. The table shows the averages of Brent crude oil prices during dramatic events and after them. It uses the same time intervals as Table 5.3. and concludes that for the Covid-19 pandemic, the average price is 44.18, compared to 67.60 for the post-Covid-19 period. Also, for the Global Financial Crisis, the average price is 58.74, compared to 76.21 for the post-Global Financial Crisis period. Additionally, while conducting my analysis, I have no reason to suspect any endogeneity problems since it is hard to detect, and the necessary information is rarely disclosed to the public. However, I recommend to future papers to include variables such as aggregate supply, global oil inventory, trading volume and aggregate demand (Gong et al., 2018) in order to conclude whether there is endogeneity in my analysis. Also, in my regression, I deal with heteroskedasticity – which says that residuals can have different variances. I test for heteroskedasticity by using the White test, which results are shown in Figures 5.3 and 5.4 for WTI and Brent crude oil prices, respectively. Both tests reject the null hypothesis of homoskedasticity. Thus, by not using the White SE, my standard errors would be too small and there would be a greater chance of making Type I error[2]. Therefore, I use White Standard

[2] Claiming statistically significant results when they should not be.

Errors (SE) to increase the efficiency and precision of my model. Finally, I will conclude whether I accept my fourth hypothesis after examining Table 5.5, which is also needed for the fifth one.

Table 5.4

Regression analysis results for both WTI Prices and Brent Prices

	WTI Prices	Brent Prices
SVI	-1.1023***	-1.3116***
Constant	83.1371***	89.5553***
Observations	4,583	4732
R^2	0.1089	0.1188

Note. Standard errors are in parentheses; * p < 0.1, ** p < 0.05, *** p < 0.01.

Table 5.5

Correlation between the SVI and crude oil prices

	WTI Prices	Brent Prices
SVI	-0.3301	-0.3447

My fifth hypothesis is based on the fourth one and says:

"If a crisis is identified, it is a profitable strategy to buy crude oil whenever the SVI volume starts to decrease."

Table 5.5. depicts the correlation between crude oil prices and the SVI. It concludes that the correlation for both WTI prices and Brent prices is negative and is -0.3301 and -0.3447 respectively. Therefore, a decrease in the SVI volume leads to an increase in crude oil prices, which is also inferred from the fourth hypothesis. Lastly, since the correlation is negative, and it is known that the SVI mostly represents the thoughts of retail investors – that are often influenced by news sources, I accept my hypothesis.

Furthermore, Table 5.5 concludes a negative correlation. Also, from the regression analysis, I have deduced that prices of crude oil and the SVI move in opposite directions. Therefore, I accept my fourth hypothesis and conclude that if there is a sharp increase in the SVI and a sharp drop in crude oil prices – then there is a dramatic event.

My last hypothesis is about forecasting and states that:

"The forecasting accuracy of crude oil prices improves when the SVI is included."

For this hypothesis, I have two models – the benchmark and the competing one. For the benchmark model, I use an ARIMA (4,1,4) while for the competing one an ARIMAX (4,1,4) is used. Lastly, the models are conducted for both WTI and Brent prices.

Table 5.7. illustrates the forecasting accuracy for WTI crude oil prices. For the competing model, the RMSE is lower, but the RMSE and MAPE are slightly higher. Additionally, the Diebold-Mariano test is insignificant with a p-value of 0.2280. This concludes that there is no significant difference in the forecasting power of the two models. Moreover, for Brent crude oil prices, the forecasting accuracy is written in Table 5.8. The RMSE, MAE, MAPE and DM are all lower for the competing model, which implies that it outperforms the benchmark one. Even though the Diebold-Mariano test is lower for the competing model, it has a p-value of 0.3373, which concludes that there is no forecasting difference between the two models. Therefore, I reject my hypothesis that the SVI improves the forecasting accuracy of crude oil prices.

Table 5.7

Forecast accuracy for WTI crude oil prices

	RMSE	MAE	MAPE	DM
Benchmark model (WTI)	1.7369	1.1177	0.01869	3.017
Competing model (includes SVI)	1.7260	1.1192	0.01871	2.979

Table 5.8

Forecast accuracy for Brent crude oil prices

	RMSE	MAE	MAPE	DM
Benchmark model (Brent)	1.4801	1.0658	0.01599	2.191
Competing model (includes SVI)	1.4779	1.0653	0.01597	2.184

Chapter 6 Conclusion

Creating a profitable strategy that uses crude oil prices is not easy since the latter is affected by many factors, some of which are economic growth, demand, trade volume, supply and investor attention. Therefore, I test a new method that includes investor attention and market overreaction as predictors of crude oil prices. The former is measured by the Google Search Volume Index (GSVI), which according to Li et al. (2015) represents data that is relatively objective compared to traditional measures. Therefore, Investor attention and more particularly the overreactions of the market which is depicted in the GSVI data have formulated my research question: *"Can retail investors profit from the overreaction hypothesis on crude oil prices?"*

In this thesis, I define the term overreaction as a drop or surge of 10% or more in WTI and Brent crude oil prices. For the former and the latter, I found 37 and 26 overreactions respectively in the period from 06/01/2004 to 20/04/2022. Furthermore, my thesis confirms that overreactions are extremely short-termed and the SVI decreases in a year after the end of the event. Opposite to stocks, it concludes that there is no reverse two-day rebound in crude oil prices and therefore an investor cannot create a profitable strategy out of this. In my thesis, I use a regression analysis that deduces that a unit increase in the SVI leads to a decrease of 1.1023 in WTI prices and 1.3116 in Brent prices. Finally, I examine the forecasting accuracy of crude oil prices by using RMSE, MAE, MAPE and DM on ARIMA (4,1,4) and ARIMAX (4,1,4) models. This leads to my conclusion, which is the same as the one of Yao and Zhang (2017), and it says that by including the SVI, the forecasting accuracy of crude oil prices does not increase. Therefore, after examining my findings, I deduce that it is extremely hard to predict crude oil prices and create a profitable strategy. Even though I have found that crude oil prices and investor attention are negatively correlated, I think that crude oil is usually highly affected by many external factors that a regular investor could not foresee. Thus, retail investors cannot profit from the overreaction hypothesis on crude oil prices, especially by only including the SVI.

The research in this thesis is useful since it tests variables that have never been used before in predicting crude oil prices. However, there are two limitations: the use of only one exogenous variable and the assumption that ARIMA (4,1,4) is the best model for forecasting crude oil prices. Therefore, for further studies, I suggest that researchers try including more variables, such as economic growth, trade volume and supply. Moreover, I recommend testing for the best model as Haque and Shaik (2021) have done and do further studies on the overreaction hypothesis on crude oil prices. Lastly, I encourage future papers to compare and assess the reliability of my findings.

REFERENCES

Borgards, O., Czudaj, R. L., & Hoang, T. H. V. (2021). Price overreactions in the commodity futures market: An intraday analysis of the Covid-19 pandemic impact. *Resources Policy*, *71*, 101966. https://doi.org/10.1016/j.resourpol.2020.101966

Bremer, M., & Sweeney, R. J. (1991). The Reversal of Large Stock-Price Decreases. *The Journal of Finance*, *46*(2), 747–754. https://doi.org/10.1111/j.1540-6261.1991.tb02684.x

Clare, A., & Thomas, S. (1995). THE OVERREACTION HYPOTHESIS AND THE UK STOCKMARKET. *Journal of Business Finance & Accounting*, *22*(7), 961–973. https://doi.org/10.1111/j.1468-5957.1995.tb00888.x

Cox, D. R., & Peterson, D. R. (1994). Stock Returns following Large One-Day Declines: Evidence on Short-Term Reversals and Longer-Term Performance. *The Journal of Finance*, *49*(1), 255–267. https://doi.org/10.1111/j.1540-6261.1994.tb04428.x

de Bondt, W. F. M., & Thaler, R. (1985). Does the Stock Market Overreact? *The Journal of Finance*, *40*(3), 793–805. https://doi.org/10.1111/j.1540-6261.1985.tb05004.x

Ellen, S. T., & Zwinkels, R. C. (2010). Oil price dynamics: A behavioral finance approach with heterogeneous agents. *Energy Economics*, *32*(6), 1427–1434. https://doi.org/10.1016/j.eneco.2010.03.003

Friedman, M. (1953). *Essays in Positive Economics*. University of Chicago Press.

Gong, X., Guan, K., Chen, L., Liu, T., & Fu, C. (2021). What drives oil prices? — A Markov switching VAR approach. *Resources Policy*, *74*, 102316. https://doi.org/10.1016/j.resourpol.2021.102316

Guo, J. F., & Ji, Q. (2013). How does market concern derived from the Internet affect oil prices? *Applied Energy*, *112*, 1536–1543. https://doi.org/10.1016/j.apenergy.2013.03.027

Han, L., Lv, Q., & Yin, L. (2017). Can investor attention predict oil prices? *Energy Economics*, *66*, 547–558. https://doi.org/10.1016/j.eneco.2017.04.018

Haque, M., & Shaik, A. (2021). Predicting Crude Oil Prices During a Pandemic: A Comparison of Arima and Garch Models. *Montenegrin Journal of Economics*, *17*(1), 197–207. https://doi.org/10.14254/1800-5845/2021.17-1.15

He, M., Zhang, Y., Wen, D., & Wang, Y. (2021). Forecasting crude oil prices: A scaled PCA approach. *Energy Economics*, *97*, 105189. https://doi.org/10.1016/j.eneco.2021.105189

Keynes, J. M. (2018). *The General Theory of Employment, Interest, and Money*. Palgrave Macmillan. https://doi.org/10.1007/978-3-319-70344-2

Li, X., Ma, J., Wang, S., & Zhang, X. (2015). How does Google search affect trader positions and crude oil prices? *Economic Modelling*, *49*, 162–171. https://doi.org/10.1016/j.econmod.2015.04.005

Loughran, T., McDonald, B., & Pragidis, I. (2019). Assimilation of oil news into prices. *International Review of Financial Analysis*, *63*, 105–118. https://doi.org/10.1016/j.irfa.2019.03.008

Mun, J. C., Vasconcellos, G. M., & Kish, R. (2000). The Contrarian/Overreaction Hypothesis. *Global Finance Journal*, *11*(1–2), 53–72. https://doi.org/10.1016/s1044-0283(00)00011-9

Tse, R. Y. (1997). An application of the ARIMA model to real-estate prices in Hong Kong. *Journal of Property Finance*, *8*(2), 152–163. https://doi.org/10.1108/09588689710167843

Vansteenkiste, I. (2011). What is Driving Oil Futures Prices? Fundamentals Versus Speculation. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1910590

Yao, T., Zhang, Y. J., & Ma, C. Q. (2017). How does investor attention affect international crude oil prices? *Applied Energy*, *205*, 336–344. https://doi.org/10.1016/j.apenergy.2017.07.131

Yegorov, Y. (2009). ABOUT ORIGINS OF FINANCIAL CRISIS 2008. *N.d.*, 1–9. https://www.researchgate.net/publication/272380285_About_Origins_of_Financial_Crisis_2008

APPENDIX

Appendix for Chapter 3

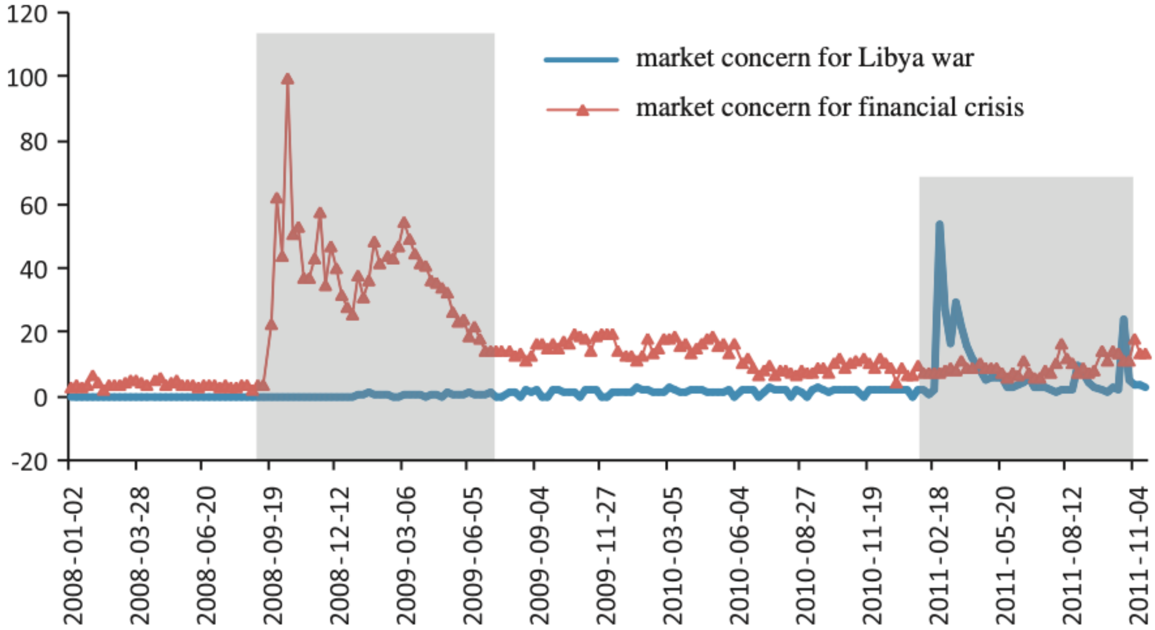


Fig. 4. Short-run market concern.

Figure 2.1. Short-run Market concern (Guo and Ji, 2013).

Appendix for Chapter 4

F505 *fx*

	A	B	C
494	04/05/2020	20.39	3.08%
495	01/05/2020	19.78	4.99%
496	30/04/2020	18.84	25.10%
497	29/04/2020	15.06	22.04%
498	28/04/2020	12.34	-3.44%
499	27/04/2020	12.78	-20.32%
500	24/04/2020	16.04	1.84%
501	23/04/2020	15.75	28.26%
502	22/04/2020	12.28	22.68%
503	21/04/2020	10.01	-126.60%
504	20/04/2020	-37.63	-305.97%
505	17/04/2020	18.27	-8.05%
506	16/04/2020	19.87	0.00%
507	15/04/2020	19.87	-1.19%
508	14/04/2020	20.11	-10.26%
509	13/04/2020	22.41	-1.54%
510	09/04/2020	22.76	-9.29%
511	08/04/2020	25.09	6.18%
512	07/04/2020	23.63	-9.39%
513	06/04/2020	26.08	-7.97%
514	03/04/2020	28.34	14.32%
515	02/04/2020	24.79	22.06%
516	01/04/2020	20.31	-0.83%
517	31/03/2020	20.48	1.94%
518	30/03/2020	20.09	-6.60%
519	27/03/2020	21.51	-4.82%
520	26/03/2020	22.6	19.01%
521	25/03/2020	18.99	-5.10%
522	24/03/2020	20.01	-1.72%
523	23/03/2020	20.36	-9.23%
524	20/03/2020	22.43	-11.06%
525	19/03/2020	25.22	23.81%
526	18/03/2020	20.37	-24.42%

Figure 4.1. Outcomes without the IF function.

C504 =IF(AND(B504<0,B505<0),IF(B504<B50

	Date	USCRWTIC Index (USD) (L1)	Average Daily Returns (USCRWT
494	04/05/2020	20.39	3.08%
495	01/05/2020	19.78	4.99%
496	30/04/2020	18.84	25.10%
497	29/04/2020	15.06	22.04%
498	28/04/2020	12.34	-3.44%
499	27/04/2020	12.78	-20.32%
500	24/04/2020	16.04	1.84%
501	23/04/2020	15.75	28.26%
502	22/04/2020	12.28	22.68%
503	21/04/2020	10.01	126.60%
504	20/04/2020	-37.63	-305.97%
505	17/04/2020	18.27	-8.05%
506	16/04/2020	19.87	0.00%
507	15/04/2020	19.87	-1.19%
508	14/04/2020	20.11	-10.26%
509	13/04/2020	22.41	-1.54%
510	09/04/2020	22.76	-9.29%
511	08/04/2020	25.09	6.18%
512	07/04/2020	23.63	-9.39%
513	06/04/2020	26.08	-7.97%
514	03/04/2020	28.34	14.32%
515	02/04/2020	24.79	22.06%
516	01/04/2020	20.31	-0.83%
517	31/03/2020	20.48	1.94%
518	30/03/2020	20.09	-6.60%
519	27/03/2020	21.51	-4.82%
520	26/03/2020	22.6	19.01%
521	25/03/2020	18.99	-5.10%
522	24/03/2020	20.01	-1.72%

Figure 4.2. Outcomes with the IF function.

The IF Function is the following:

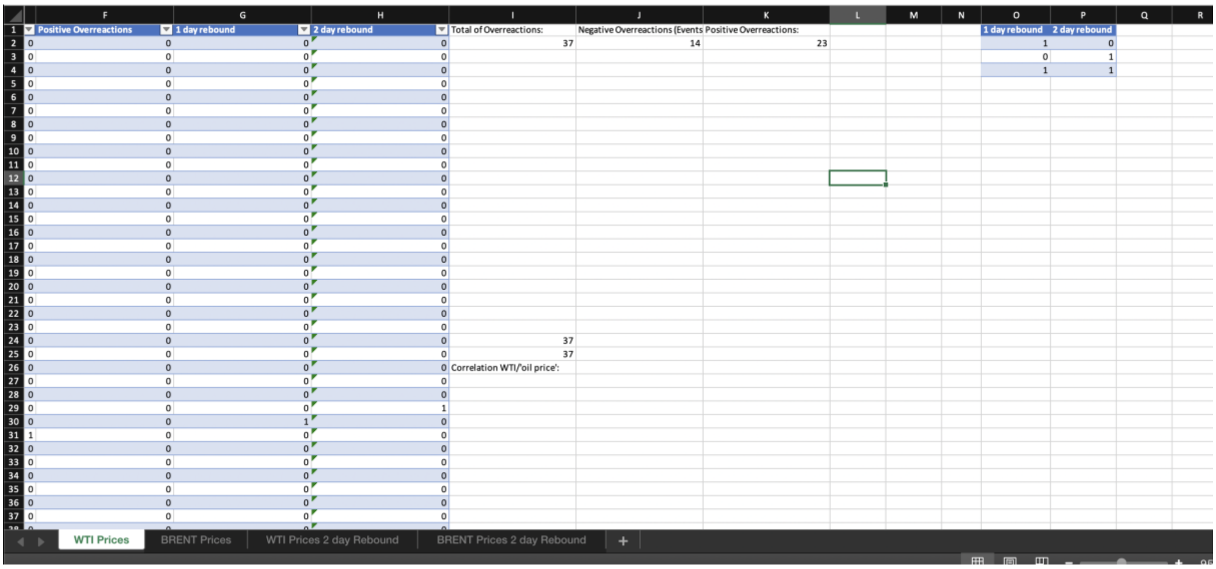
=IF(AND(B504<0, B505<0),IF(B504<B505,(-((B504/B505)-1)),((B504/B505))-1),IF(OR(B504<0,B505<0),IF(B504>B505,-((B504/B505)-1),((B504/B505)-1)),((B504/B505)-1)))

If there was a module instead of this function, I would have gotten only positive returns, therefore I cannot use a module. That's why I decided to create a universal function that can be used for every single cell and that can smartly adjust the returns.

To find the number of overreactions I can use one of the two methods:

The formula to find the negative overreactions in Excel: =IF(OR(C2<= -10%),1,0) → This equation gives us a 1 if the cell, namely C2 is smaller or equal to -10%, and otherwise. After, I have found the negative overreactions I can find the positive ones with a similar formula: =IF(OR(C2>=10%),1,0). Then I can sum up the number of negative and positive events with the function '=SUM' that uses arrays for convenience.

The other method is just to use the following formula: =IF(OR(C2>=10%,C2<= -10%),1,0). It checks for cells that have either an average daily return smaller than -10% or greater than 10%. If any of those conditions is satisfied, then we get a return of 1. Later, I can sum up those returns by using =SUM.



		F	G	H	I	J	K	L	M	N	O	P	Q	R
1		Positive Overreactions	1 day rebound	2 day rebound	Total of Overreactions:	Negative Overreactions (Events	Positive Overreactions:				1 day rebound	2 day rebound		
2	0	0	0	0	37	14	23				1	0		
3	0	0	0	0							0	1		
4	0	0	0	0							1	1		
5	0	0	0	0										
6	0	0	0	0										
7	0	0	0	0										
8	0	0	0	0										
9	0	0	0	0										
10	0	0	0	0										
11	0	0	0	0										
12	0	0	0	0										
13	0	0	0	0										
14	0	0	0	0										
15	0	0	0	0										
16	0	0	0	0										
17	0	0	0	0										
18	0	0	0	0										
19	0	0	0	0										
20	0	0	0	0										
21	0	0	0	0										
22	0	0	0	0										
23	0	0	0	0										
24	0	0	0	0	37									
25	0	0	0	0	37									
26	0	0	0	0	Correlation WTI/'oil price':									
27	0	0	0	0										
28	0	0	0	0										
29	0	0	0	1										
30	0	0	1	0										
31	1	0	0	0										
32	0	0	0	0										
33	0	0	0	0										
34	0	0	0	0										
35	0	0	0	0										
36	0	0	0	0										
37	0	0	0	0										

Figure 4.3. Advanced Filter.

Figure 4.3 depicts the advanced filter option. My list of criteria is the small table in columns "O" and "P". Furthermore, the columns "G" and "H" are my list ranges which represent the 1 day rebound and 2 day rebound respectively.

Appendix for Chapter 5

```
. reg WTIPrice crudeoilWTI

      Source |       SS           df       MS      Number of obs   =     4,583
-------------+----------------------------------   F(1, 4581)      =    560.08
       Model |  257937.411         1  257937.411   Prob > F        =    0.0000
    Residual |  2109707.59     4,581  460.534292   R-squared       =    0.1089
-------------+----------------------------------   Adj R-squared   =    0.1087
       Total |     2367645     4,582  516.727412   Root MSE        =     21.46

------------------------------------------------------------------------------
    WTIPrice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 crudeoilWTI |  -1.102318    .046578   -23.67   0.000    -1.193634   -1.011003
       _cons |   83.13712   .6826247   121.79   0.000     81.79885     84.4754
------------------------------------------------------------------------------
```

Figure 5.1. STATA Output for the regression on WTI crude oil prices.

```
. reg BRENTPrice crudeoil2

      Source |       SS           df       MS      Number of obs   =     4,732
-------------+----------------------------------   F(1, 4730)      =    637.63
       Model |   374124.29         1   374124.29   Prob > F        =    0.0000
    Residual |   2775283.4     4,730  586.740676   R-squared       =    0.1188
-------------+----------------------------------   Adj R-squared   =    0.1186
       Total |  3149407.69     4,731  665.695981   Root MSE        =    24.223

------------------------------------------------------------------------------
  BRENTPrice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   crudeoil2 |  -1.311572   .0519406   -25.25   0.000      -1.4134   -1.209744
       _cons |   89.55534   .7563589   118.40   0.000     88.07252    91.03815
------------------------------------------------------------------------------
```

Figure 5.2. STATA Output for the regression on Brent crude oil prices.

Table 5.6

Averages of Brent crude oil price during dramatic events and after them

	Covid-19 Pandemic (20/11/2019 – 19/11/ 2020)	Post Covid-19 (20/11/2020 – 20/11/2021)	Global Financial Crisis (15/09/2008-15/09/2009)	Post Global Financial Crisis (16/09/2009 – 16/09/2010)
Crude oil	44.18	67.60	58.74	76.21

```
White's general test statistic :  228.7562  Chi-sq( 2)  P-value =  2.1e-50
```

Figure 5.3. White test for WTI Prices.

```
White's general test statistic :  291.5301  Chi-sq( 2)  P-value =  5.0e-64
```

Figure 5.4. White test for Brent Prices.

```
Diebold-Mariano forecast comparison test for actual : WTIPrice
Competing forecasts:  hat versus hatex
Criterion: MSE over 3599 observations
Maxlag = 29 chosen by Schwert criterion   Kernel : uniform

Series                      MSE
------------------------------
hat                         3.017
hatex                       2.979
Difference                 .03769

By this criterion, hatex is the better forecast
H0: Forecast accuracy is equal.
S(1) =      1.205   p-value = 0.2280
```

Figure 5.5. Diebold-Mariano test for WTI prices.

```
Diebold-Mariano forecast comparison test for actual : BRENTPrice
Competing forecasts:  hat versus hatex
Criterion: MSE over 3767 observations
Maxlag = 29 chosen by Schwert criterion   Kernel : uniform

Series                 MSE
------------------------------
hat                    2.191
hatex                  2.184
Difference           .006329

By this criterion, hatex is the better forecast
H0: Forecast accuracy is equal.
S(1) =     .9594  p-value = 0.3373
```

Figure 5.6. Diebold-Mariano test for Brent prices.