School of Economics, Finance and Property

The Relationship between Crude Oil Price, Exploration, Production, and Uncertainty

Esti Tri Widyastuti

This thesis is presented for the collaborative Degree of
Doctor of Philosophy
of

University of Aberdeen

and

Curtin University

November 2022

Declaration

This thesis has been written as part of the Doctoral Programme at the University of Aberdeen in collaboration with Curtin University. All quotations have been distinguished by quotation marks, and all sources of information have been specifically acknowledged. This thesis contains no material previously published by anyone except where due acknowledgement has been made.

Signed:

Date: November 11, 2022

Abstract

The many extreme oil price episodes that have occurred in the past have sparked an ongoing debate as to the main drivers of oil price fluctuations and their effects on the economy. The literature has widely discussed the important role played by fundamental supply and demand as a key contributor to oil price fluctuations. This thesis, however, is concerned with the opposite effect — that of the crude oil price effect on oil supply and demand. There is a feedback loop between crude oil supply, demand, uncertainty, and oil prices. While demand is strongly influenced by global economic activity, supply is more complicated, being the result of exploration activities in the past. These, again, are strongly influenced by oil prices. In addition, there is an uncertainty issue that affects oil prices. Uncertainty is captured by the public interest in unexpected circumstances that disrupt supply and demand, leading to oil price fluctuation. However, little attention has been paid by the existing literature to exploration activity and the specific uncertainty measures for the global oil market.

This thesis addresses the relationship between crude oil price, exploration, production, and uncertainty. Chapter 2 emphasises the importance of exploration activity in the oil market, taking the case study of a mature petroleum province, the Norwegian Continental Shelf, from 1966 to 2019. Chapter 2 contributes to the academic literature in two ways. First, by providing a toolbox for applying a Monte Carlo simulation to capture the effect of exploration activity. Second, empirically analysing the relationship between exploration

activity and crude oil price. The simulation results illustrate that the more frequent the oil discovery, the more exploration activity there will be, as well as a shorter waiting time between oil discoveries. A longer waiting time between discoveries is necessary for finding a large discovery, which is less frequently found than small discoveries. By applying an Autoregressive Distributed Lag (ARDL) model, this study finds a long-run relationship between crude oil price, exploratory effort, and efficiency. Oil price does not affect exploratory efficiency in the short term; only exploratory effort does that. However, the long-run relationship between crude oil price and exploration influences oil supply in the future. A structural break causes negative effects on exploratory efficiency, and the model with a break in 2008 confirms the long-run relationship between oil price, exploratory effort, and efficiency.

The discussion of the oil supply side is presented in Chapter 3. Chapter 3 contributes to the literature by investigating the asymmetric effect of crude oil price shocks, looking at whether oil price increases and decreases have different effects on global oil supply and demand. Applying the basic ARDL and non-linear ARDL for time series data comprising world crude oil production, the Baltic Dry Index (BDI), and real oil prices from January 1985 to December 2019, this chapter finds a long-run equilibrium relationship between crude oil price, supply, and demand. The result suggests that oil price increases have a stronger effect on world oil production and the BDI than on oil price decreases. Adding a structural break in January 2009 confirms the long-run relationship and the stronger impact of oil price increases on global demand.

In addition, there is an uncertainty component that cannot be explained by supply and demand. This lacuna is explored in Chapter 4, contributing to the literature on uncertainty in two ways. First, by constructing a newly proposed Google Trends-based Uncertainty (GTU) index specific to the global oil market uncertainty measure. Second, empirically analyse the relationship between uncertainty, exploration, and crude oil price. The study finds that the

GTU index, which measures intensified public interest at times of uncertainty, is closely related to crude oil price movements, particularly when oil prices drop. The GTU index can capture unexpected events associated with marked spikes or drops in the crude oil price, as occurred during the global financial crisis in 2008-09, the oil price drop in 2014-2016, and Coronavirus in 2020. The chapter's empirical analysis applying time-varying Granger causality and Vector Autoregressive (VAR) for the GTU index, oil rig counts (disaggregating between the world and North America), and the crude oil price from January 2004 to April 2020 suggests that uncertainty indices Granger cause rig counts and oil price and bidirectional causality between oil price and rig counts in most of the uncertainty indices. Uncertainty negatively affects oil prices and rig counts. Further, that oil price negatively affects uncertainty and positively affects rig counts.

This study undertakes an academic exercise to understand exploration activity, which can be applied to other petroleum regions. Furthermore, understanding the effect of exploration activity and how it is affected by oil price movements reveals insights that will be of assistance to the preliminary study of exploration in the frontier areas. The asymmetric effect analysis offers implications for the supply side by suggesting that it is necessary to boost world oil production in the long term to overcome extreme oil price increases and to ensure the security of supply to meet the strong global demand. The newly proposed GTU index provides a specific measure of uncertainty in the global oil market, is easy to construct, and results in a distinct index that captures how public interest, is of components that cannot be captured by supply and demand, contributes to oil price shocks.

Acknowledgements

First, I would like to express the highest gratitude to Allah Almighty, the Most Gracious and the Most Merciful. Alhamdulillaah, praise be to Allah Almighty for the abundant and continuous blessings, who is the light in every stage of my life and gives me the strength to complete this PhD thesis.

I would like to acknowledge the Aberdeen – Curtin Alliance for the funding support, which allows me to gain experience by studying at the University of Aberdeen, UK and Curtin University, Australia. My sincerest gratitude to my supervisors, Dr Marc Gronwald and Dr Hiroaki Suenaga, for the excellent research support, continuous guidance, and constructive feedback throughout my PhD journey, and Dr Nikolaos Vlassis, for being tremendously supportive and always motivating me, particularly during the most difficult time in my last stage of the study. I would also like to express my deep gratitude to the examiners, Professor Euan Phimister (University of Aberdeen), Dr Lei Pan (Curtin University), Dr Xiaoyi Mu (University of Dundee), and Dr Raymond Li (University of Canberra), for their advice and feedback on my thesis and great experience and discussion during my Viva Voce.

Most importantly, my biggest thanks to my late parents, Papa Sugeng Sudharto and Mama Endang Sugiastuti, for their unconditional love, support, encouragement, and every single thing, especially my beloved late mama, who is always with me and my soul, in the past, current, and future. There is no such word I can describe to say how much I dedicate everything to you, and

I hope we will have the best time later when we reunite with dad and sister, Endah Dwi Wahyuni. My brother, Andy Sucahyo, let's do our best to make mama papa proud.

Finally, I would also like to thank my great communities, special thanks to 'Pongah' members (Sis Sri, Uche, Davina, Emmi, Maisara, and Detty) for being with me in my happiest and lowest stage. Last but not least, I thank PPI Aberdeen, Sakinah's sisters (Mama Nur, Umi Indah, Mommy Loli, Mommy Vidy, and Mama Stela), Aberdeen - Curtin Alliance students, and all members and alumni of Edward Wright Annexe F04, F02, and F03, for all the cherished fun moments.

Contents

1 Introduction **1**

2 Exploration activity and the crude oil price: evidence from the mature petroleum province **21**

2.1 Introduction 21

2.2 The literature on exploratory effort and efficiency 31

2.3 Oil discovery data and descriptive statistics 38

2.3.1 Norwegian Continental Shelf (NCS) discovery sequences 38

2.3.2 Exploratory effort, efficiency, and crude oil price 40

2.4 Methodology 44

2.4.1 Monte Carlo simulation 44

2.4.1.1 Independence test 45

2.4.1.2 Statistical measures 46

2.4.1.3 The candidate of continuous probability distributions and parameter estimates 47

2.4.1.4 Model selection by information criteria 50

2.4.1.5 The goodness-of-fit statistics 51

2.4.1.6 Autoregressive Moving Average (ARMA) forecast model 52

2.4.1.7 Evaluation test 53
2.4.2 Empirical framework 53
2.4.2.1 Unit root test 53
2.4.2.2 Structural break test 53
2.4.2.3 Autoregressive Distributed Lag (ARDL) 55
2.4.2.4 Autoregressive Distributed Lags in Error-Correction form (ARDL-EC) and ARDL with a break . . . 56
2.5 Distribution type of NCS oil discovery size and time between discoveries 59
2.6 Simulation results and analysis 63
2.6.1 Simulation of discovery waiting time and size in a mature petroleum province 63
2.6.2 Empirical results of exploratory effort, efficiency, and crude oil price 73
2.6.2.1 The empirical ARDL model with a structural break 79
2.7 Conclusion 86

3 World oil production, global demand, and the crude oil price: evidence of the asymmetric effect 89

3.1 Introduction 89
3.2 The literature on crude oil price and production and contribution of the study 96
3.2.1 The extension of global crude oil market model 97
3.2.1.1 The role of speculation 97
3.2.1.2 The use of disaggregated data 98

3.2.1.3 Reduced-form VAR implementation 100
3.2.1.4 The difference in the time series data treatment 100
3.2.1.5 The inclusion of other factors in the modelling . 102
3.2.2 The contributions to the existing literature 103
3.3 Global oil market data and descriptive analysis 105
3.3.1 Description of the data 105
3.3.2 Descriptive analysis of data 108
3.3.3 Unit root tests 109
3.3.3.1 Enders and Lee (2012) test 110
3.3.3.2 Unit root tests result 112
3.4 Methodology 114
3.4.1 The ARDL and ARDL in Error-Correction form (ARDL-EC) 115
3.4.2 Nonlinear Autoregressive Distributed Lags (NARDL) . . 118
3.5 Empirical result: the relationship between world oil production, demand, and crude oil price 121
3.5.1 The basic ARDL results 121
3.5.1.1 ARDL in error correction interpretation 125
3.5.2 The asymmetric ARDL (NARDL) results 132
3.5.2.1 NARDL in error correction interpretation . . . 136
3.6 Conclusion 152

4 Global oil market uncertainty, oil exploration, and crude oil price: an application of Google Trends 155

4.1 Introduction 155

4.2 The literature on uncertainty measures and contribution of the study 164
4.2.1 Uncertainty and crude oil price 172
4.2.2 Uncertainty and economic activity 175
4.2.3 Contribution of the study 177
4.3 Uncertainty measures and oil market data 182
4.4 Google Trends-based Uncertainty (GTU) indices: construction, comparison, and discussion 186
4.4.1 The newly proposed GTU indices 187
4.4.1.1 GTU oil price 190
4.4.1.2 GTU oil investment, GTU oil supply, and GTU oil demand 192
4.4.1.3 GTU oil market specific 196
4.4.2 Descriptive statistics 200
4.4.3 Correlation between GTU indices and other uncertainty benchmark measures 203
4.5 Methodology 210
4.5.1 VAR framework 211
4.5.1.1 Granger Causality test 212
4.5.1.2 Impulse response function 214
4.5.1.3 Variance decomposition 215
4.6 Empirical Results and Analysis 217
4.6.1 Unit Root Test 217
4.6.2 Time-varying Granger causality test 221

4.6.3 VAR and Impulse responses to rig counts, oil price, and uncertainty shocks . . . 225
4.6.4 Forecast error variance decomposition of oil market to uncertainty shock . . . 231
4.7 Conclusion . . . 234

5 Conclusion 237

A Exploration activity and the crude oil price: evidence from the mature petroleum province 247
A.1 R code for Norwegian oil discovery simulation . . . 247
A.2 Robustness test . . . 274

B Global oil market uncertainty, oil exploration, and crude oil price: An application of Google Trends 276
B.1 The optimal lag length selection . . . 276
B.2 Unit root test . . . 280
B.3 VAR estimates . . . 280
B.4 Autocorrelation test . . . 297
B.5 Robustness test . . . 298
B.5.1 Reordering variables . . . 298
B.5.2 Brent oil price . . . 303

List of Tables

2.1 Descriptive statistics of NCS discovery data 40

2.2 Pairwise correlation between waiting time, other exploratory efforts, and crude oil price 42

2.3 ARDL model deterministic cases 56

2.4 Goodness-of-fit criteria for discovery size and time between discoveries distribution 62

2.5 BDS Independence test 64

2.6 Decile of simulated discovery size 65

2.7 Decile of simulated discovery waiting time 65

2.8 Decile of simulated discovery dates 66

2.9 Mean Absolute Error (MAE) comparison of the Monte Carlo simulation and the ARMA(1,1) estimation 68

2.10 Unit Root Test for NCS discovery, exploratory effort, and crude oil price variables 75

2.11 ARDL estimation for exploratory effort, efficiency, and crude oil price 75

2.12 ARDL estimation in error correction form for exploratory effort, efficiency, and crude oil price 76

2.13 Structural break test result for waiting time between discovery, discovery size, exploration well, and crude oil price 80

2.14 ARDL estimation with a structural break for exploratory effort, efficiency, and crude oil price 81

2.15 ARDL estimation in error correction form with a structural break in 2008 for exploratory effort, efficiency, and crude oil price 84

3.1 Descriptive statistics of crude oil market data 108

3.2 Unit root test for world crude oil production, BDI, and crude oil price 113

3.3 ARDL estimation for world oil production, BDI, and crude oil price 122

3.4 Structural break test result for world crude oil production, BDI, and crude oil price 123

3.5 ARDL with a structural break model for world oil production, BDI, and crude oil price 124

3.6 ARDL estimation in error correction form for world oil production, BDI, and crude oil price 126

3.7 ARDL estimation with a structural break in error correction form for world oil production, BDI, and crude oil price 131

3.8 NARDL estimation for world oil production, BDI, and crude oil price 133

3.9 NARDL estimation with a structural break for world oil production, BDI, and crude oil price 135

3.10 NARDL estimation in error-correction form for world oil production, BDI, and crude oil price 138

3.11 NARDL estimation with a structural break in error-correction form for world oil production, BDI, and crude oil price 146

4.1 Base Search Terms for GTU Oil investment, GTU Oil Supply, and GTU Oil Demand 189

4.2 Descriptive Statistics (level) 202

4.3 Pairwise correlation between uncertainty indices and oil market variables (in level) 208

4.4 Unit root test for uncertainty and oil market variables 219

4.5 Structural break test result 220

4.6 Time Varying Granger Causality test between uncertainty and global oil market 222

4.7 Root Mean Squared Error (RMSE) of the predicted rig counts, crude oil price, and uncertainty 225

4.8 Forecast Error Variance Decomposition 232

A.1 ARDL estimation with a structural break in error correction form for exploratory effort, efficiency, and adjusted PPP real oil price 275

B.1 VAR optimal lag length based on information criteria 279

B.2 Narayan and Popp (2010) Unit Root Test for Uncertainty and Oil Market Variables 280

B.3 VAR estimates for world exploration activity 281

B.4 VAR estimates for North America exploration activity 289

B.5 Autocorrelation test 298

List of Figures

1.1 Overview of the thesis 6

1.2 Time series of crude oil price, rig counts, and Google Trends . . 7

2.1 Global petroleum production, consumption, and crude oil price . 22

2.2 Global discoveries for 2021 on course to lowest in decades 24

2.3 Accumulated resources on the Norwegian continental shelf . . . 28

2.4 Norwegian Continental Shelf oil discoveries 1967-2019 39

2.5 Norwegian Continental Shelf oil discovery waiting time 39

2.6 The Norwegian oil discovery waiting time and crude oil price . . 43

2.7 The Norwegian oil discovery waiting time and exploration well . 43

2.8 Skewness and Kurtosis of NCS oil discovery size 59

2.9 Skewness and Kurtosis of NCS time between discoveries 60

2.10 Empirical and theoretical NCS oil discovery size distribution . . 61

2.11 Empirical and theoretical NCS time between discoveries distribution 61

2.12 Density and Empirical Cumulative Distribution Function of simulated size and waiting time 67

2.13 Simulated NCS discovery size and year 70

2.14 Simulated discovery size and year (low mean of discovery size and rate) . . . 71

2.15 Simulated discovery size and year (high mean of discovery size and rate) . . . 72

2.16 The simulated Norwegian oil discovery waiting time and crude oil price . . . 74

2.17 The simulated Norwegian oil discovery waiting time and oil exploration well . . . 74

3.1 World crude oil production . . . 105

3.2 Baltic Dry Index . . . 106

3.3 Crude oil price . . . 108

3.4 Oil price and supply adjustment . . . 144

3.5 Oil price and demand adjustment . . . 144

3.6 Oil supply and demand adjustment . . . 145

3.7 World oil supply, global demand, and oil price dynamic multipliers 150

3.8 World oil supply, global demand, and oil price dynamic multipliers with a structural break . . . 151

4.1 The Uncertainty Benchmark Measures for full sample period . . 165

4.2 The Uncertainty benchmark measures for sample period: May 2007-April 2020 . . . 166

4.3 World oil rig counts, North America oil rig counts, and crude oil prices . . . 185

4.4 Search term 'petroleum' and 'oil' comparison in Google Trends . 188

4.5 GTU oil price index . . . 191

4.6 GTU oil price index and global oil market variables . . . 191

4.7 GTU oil investment . . . 194

4.8 GTU oil investment and the global oil market variables . . . 194

4.9 GTU oil demand . . . 196

4.10 GTU oil demand and the global oil market variables . . . 197

4.11 GTU oil supply . . . 197

4.12 GTU oil supply and the global oil market variables . . . 197

4.13 GTU oil market specific index . . . 198

4.14 GTU oil market specific index and global oil market variables . 198

4.15 GTU indices for oil price, oil investment, oil supply, oil demand, and oil market specific . . . 199

4.16 The GTU Oil Market Specific Indices and Other Existing Uncertainty Measures . . . 210

4.17 Response to Cholesky one-standard deviation shock in world oil exploration model . . . 227

4.18 Response to Cholesky one-standard deviation shock in North America oil exploration model . . . 228

B.1 Uncertainty indices (log differences) . . . 277

B.2 Oil market variables (log differences) . . . 278

B.3 World oil exploration activity responses to various uncertainty shocks (reordering: uncertainty - rig count - oil price) . . . 299

B.4 North America oil exploration activity responses to various uncertainty shocks (reordering: uncertainty - rig count - oil price) . 300

B.5 Crude oil price responses to various uncertainty shocks (reordering: uncertainty - rig count - oil price) - world exploration activity model . . . 301

B.6 Crude oil price responses to various uncertainty shocks (reordering: uncertainty - rig count - oil price) - North America exploration activity model 302

B.7 Rig count response to GTU oil market specific shock (Brent price model) 303

B.8 Rig count response to Brent price shock 303

B.9 Brent price response to GTU oil market specific shock 304

Chapter 1

Introduction

Oil is the primary energy input for industry and its price is historically subject to fluctuation. These two factors motivate the large body of literature that attempts to analyse the relationship over time between crude oil prices and the macroeconomy. Early literature by Hamilton (1983, 1985) primarily identifies oil price shocks as contributing to most of the major recessions experienced by the United States (US). This view considers oil price increase to be an exogenous shock caused by supply disruption arising from geopolitical conflicts, such as the Suez Crisis (January 1957 – February 1957), OPEC embargo (November 1973 – February 1974), and the first Persian Gulf War (August 1990 – October 1990).

The more recent literature presents new insight, which is that oil price shock should be viewed as an endogenous shock due to strong demand growth, particularly in Asia, and market concern about future oil supply (Kilian, 2009). These two demand shocks cause an increase in oil price spikes that is more persistent and substantial than that which is caused by supply disruption. The difference between these two demand shocks is that market concern shock causes an immediate effect on the real oil price, whereas the global demand shock has a delayed effect. The argument is that the fluctuation in demand for global industrial commodities and uncertainty about the future oil shortages are the major contributors to oil price shocks. Oil price episodes caused by such demand and uncertainty shocks have occurred after the Iranian revolution (1979), the Venezuelan oil strike and second Gulf War (2003), the Global Financial Crisis (2008), and the oil price decline between 2014 and 2016 (Kilian and Murphy, 2014; Baumeister and Kilian, 2016*b*). Gronwald (2016) explains the temporary explosions in oil price in 1990–91, 2005–06, and 2008–09 as the result of the change in fundamental supply and demand, and the low price elasticities of supply and demand. Revisiting the role of supply and demand, Baumeister and Hamilton (2019) find that supply shock has a larger effect on the decline in economic activity than had been identified in Kilian (2009); Kilian and Murphy (2014) which argue that supply shock plays a minor role in explaining oil price fluctuation.

Aside from the supply and demand factors, market expectation is also historically believed to be an essential factor contributing to oil price increase. Kilian (2009) calls this type of shock a ‘speculative demand shock’ since it is one that cannot be explained by supply and demand shocks. This factor is associated with uncertainty and it plays an important role, even more so once crude oil became a global commodity trading instrument. The underlying supply and demand are insufficient for explaining this type of shock, as it depends on unanticipated market behaviour. Uncertainty shock is of concern to the oil industry’s upstream businesses, policymakers, traders, and investors due to the complications of measuring it.

Supply shock is the outcome of fluctuations in crude oil exploration and production activities. Exploration activity determines future production, making it a critical stage and one that requires consideration ahead of oil production. However, recent research mainly pertains to oil production, with only a few studies investigating the relationship between exploration and crude oil price. These include Toews and Naumov (2015), who emphasise the strong positive significant relationship between global drilling activity and the oil price. That being said, the existing literature on oil exploration mainly focuses on predicting the remaining US oil reserves (such as work done by Smith (1980); Cleveland and Kaufmann (1991); Laherrere (2002)), and is thus comes from the fields of petroleum geology, engineering, and other statistics-based work. The oil exploration literature contains comprehensive analyses of the mature petroleum provinces (e.g., Kemp and Kasim (2003) and Mohn and Osmundsen (2008)). Kemp and Kasim (2003) incorporate geological factors with economics in their UK Continental Shelf simultaneous equations model, while Mohn and Osmundsen (2008) propose an econometric model for the exploration and appraisal drilling activity in the Norwegian Continental Shelf. Mohn and Osmundsen (2008) conclude that oil price has a long-term effect on exploratory effort, and that licensing policy and the historical success discovery rate exert a short-term effect on exploration activity. Another study by Ewing (2017) applies US data to an examination of oil price fluctuation on the newly proved reserves. The lack of studies in the economic literature combining economics with geological components of oil exploration motivates this thesis to examine exploratory effort and efficiency in response to crude oil price fluctuation.

The large body of existing literature discusses mainly the relationship between fundamental oil supply, oil demand, and crude oil price in a short-term and linear relationship. The earlier literature applies reduced form vector autoregression to predict the variable of interest, with a model that captures feedback effects to show how the current and past values of the variable of interest affect each other (Enders, 2015). Sims (1980) applies the model in the macroeconomy to predict real fluctuation by variances in money stock, industrial production, and wholesale price innovations. Hamilton (1983) adds the role of oil to the macroeconomy variables previously discussed in Sims (1980), and finds that GNP growth decline (leading to recession) followed oil price increases between 1948 and 1972. Later, the literature in the oil market model commonly applies the structural vector autoregression model (see (Kilian, 2009; Ratti and Vespignani, 2015). However, this model requires all variables to be stationary and has the limitation of not being capable of analysing the long-run relationship.

Certain works do consider the analysis of the long-run relationship, for instance, Kolodzeij and Kaufmann (2014). The cointegrated vector autoregression model is applied to disentangle the long-run relationship from the short-run. However, the cointegrated model does not allow a mix of stationary and non-stationary variables, such as often occurs in time series, and it is thus most suitable for variables integrated with the same order. Furthermore, most studies assume that the oil price shocks have symmetric effects on supply and demand. However, positive or negative oil price shocks may differently affect the global oil market variables. A positive shock may have a stronger or weaker effect on the supply and demand; hence, an asymmetric effect. The recent literature discusses the asymmetry application in the macroeconomic field, examining, say, how the macroeconomic variables respond to a spike or drop in oil price and the effects of economic expansion and downturn, but there is less discussion specific to oil supply and demand response. It is also essential to understand the long-run relationship between oil price, supply, and demand because the deviation that causes disequilibrium in the short run is corrected in the long run. Distinguishing between short- and long-run relationships helps the decision maker and potential investor make short-, medium-, and long-term strategies for the unexpected oil price fluctuation.

The terms 'crude oil price' and 'uncertainty' are closely linked. The existing studies interpret uncertainty as relating to market expectation, public interest, or sentiment indicators about future oil prices. Pindyck (1980) defines uncertainty as the unknown random fluctuation of the future value of demand and available reserves with known current demand and reserves. Kellogg (2014) interprets uncertainty as the market expectation of future oil price volatility in a study that analyses the effect of uncertainty on exploration investment behaviour. The author finds that when the expected volatility of future oil price increases, drilling activities in Texas decline. Another interpretation of uncertainty as a market concern is defined by Alquist and Kilian (2010) and Kilian and Murphy (2014). The authors define uncertainty as to market behaviour where crude oil is purchased for future consumption as the result of an anticipated shortfall in future oil supply.

The focus of the existing literature on uncertainty measures macroeconomic, financial, and policy indicators, and there is no specific index to measure global oil market uncertainty by linking oil exploration, production, and oil price. The current measure of uncertainty is commonly approached by using macroeconomic indicators and forecast-based volatility (Jurado et al., 2015; Rossi and Sekhposyan, 2015), crude oil inventories (Kilian and Murphy, 2014), and the traditional near-term volatility indices. Most of these uncertainty indices are not directly linked to unexpected circumstances affecting oil price movements, being mainly constructed from macroeconomic variables and tailored to disaggregate country-based data. There have been recent attempts to connect market expectation and unexpected events by approaching uncertainty via media-based (newspaper or internet) query volume. Most research focuses on proposing the macroeconomic uncertainty index, with only a few studies suggesting a direct uncertainty index for the global oil market. Such studies include the work done by Guo and Ji (2013), which employs Google search volume to measure market concern about oil price and demand components.

This thesis has three main discussions of the relationship between exploration, production, uncertainty, and crude oil price shocks. It thus addresses the shortcomings in the global oil market literature, as illustrated in Figures 1.1 and 1.2. Figure 1.1 shows there is a feedback loop between crude oil supply, demand, uncertainty, and oil prices. In the short run, supply and demand determine oil prices. While demand is strongly influenced by global economic

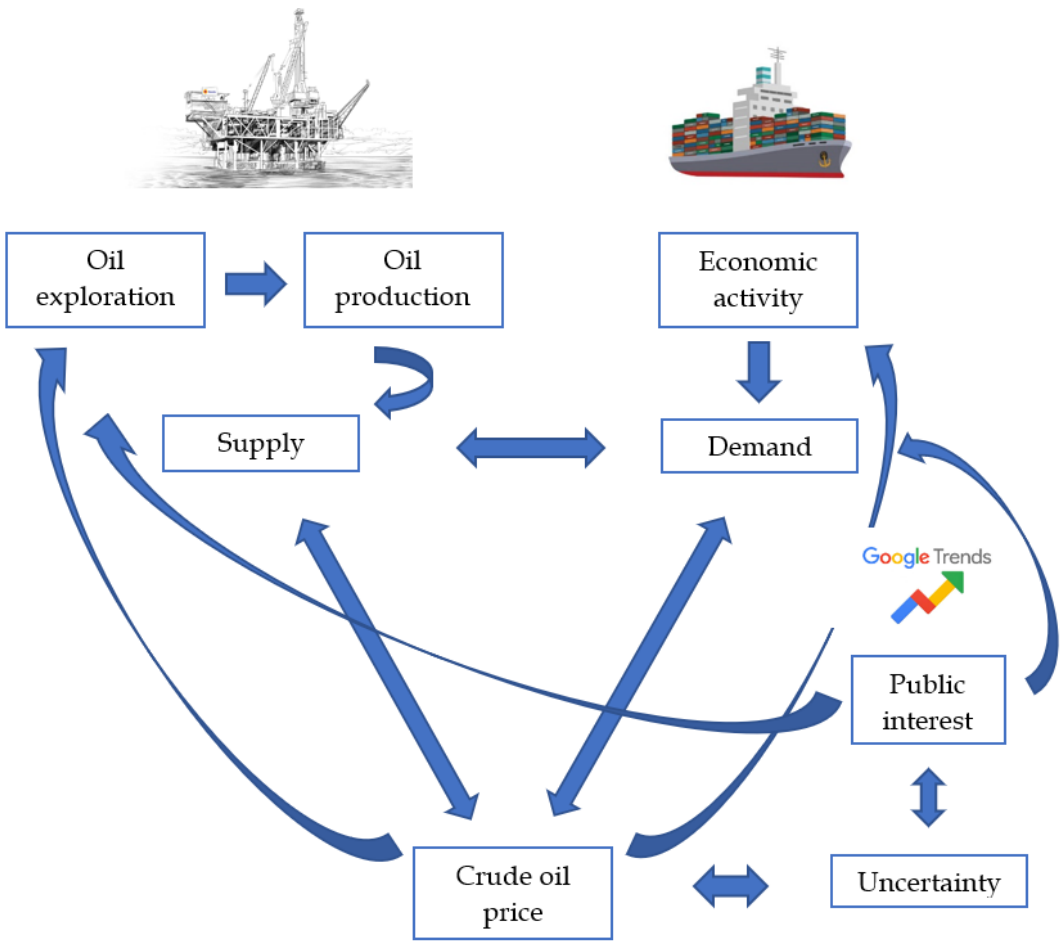


Figure 1.1: Overview of the thesis

activity, supply is more complicated, being a result of exploration activities in the past. These, again, are strongly influenced by oil prices. Thus, in the long run, the direction of oil supply influences changes. In addition, there is an uncertainty issue that affects oil prices. The uncertainty is captured by public interest in the unexpected circumstances that disrupt supply and demand, leading to oil price fluctuation.

There are a number of research questions addressed in this thesis. The first two are: how does exploration activity behaviour over time? How does crude oil price fluctuation affect exploration? These two research questions are discussed in Chapter 2, which emphasises the role of oil exploration. Chapter 3 asks what is the relationship between crude oil supply, demand, and oil prices? How do positive and negative shocks in the crude oil price affect supply and demand differently? These research questions focus on crude oil production, measuring oil supply in the long run and the asymmetric effects of oil price shocks on supply and demand. Finally, Chapter 4 discusses uncertainty measures in the global oil market and the relationship with the crude oil price by asking three research questions. First, how can the Google Trends-based Uncertainty index be used as a measure of global oil market uncertainty and what difference might the construction of a more refined GTU index make to the literature and its findings? Second, what precisely do these GTU indices measure, and how do their measurements compare to those of the existing indices? Third, what is the relationship between uncertainty, crude oil price, and exploration?

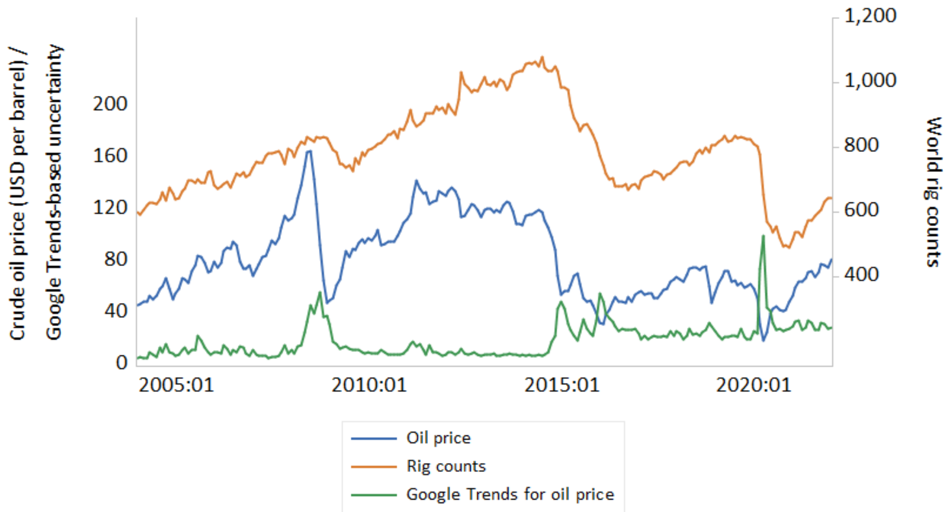


Figure 1.2: Time series of crude oil price, rig counts, and Google Trends

The starting point of this thesis is Chapter 2, which takes oil prices as given in focusing on how they affect oil exploration. A Monte Carlo simulation exercise is carried out to capture the effect over time of exploration activity in a mature petroleum province. An empirical estimation addresses the relationship between oil exploration and oil price. While the interaction between oil exploration and production will be left for future research, it will be acknowledged that there are some lags between exploration and the actual production.

The discussion of the supply-side captured by world crude oil production is presented in Chapter 3. Chapter 3 extends the literature by applying the asymmetric model for world oil production, global demand, and crude oil price relationship. The econometric framework is carefully chosen to distinguish long-run equilibrium from the short-run dynamic relationship.

The role of uncertainty in capturing other components that cannot be explained by fundamental supply and demand is discussed in Chapter 4. Chapter 4 approaches global oil market uncertainty through public interest by implementing the newly proposed Google Trends-based Uncertainty (GTU) index. The first section of Chapter 4 presents the construction of the GTU index and offers a comparison between the predictability of GTU and the benchmark indices on oil price and exploration. The second section analyses the relationship between uncertainty, oil price, and exploration.

The visual time series plot illustrated in Figure 1.2 summarises the entire thesis. The plot shows a relationship between crude oil price and world rig counts, with the latter measuring exploration activity. The figure indicates a strong co-movement between these two variables; in particular, when crude oil prices drop, rig counts also dramatically decline, as can be seen in the cases of the post-Global Financial Crisis in 2009, the oil price decline in 2016, and the Coronavirus outbreak in 2020. In addition, there is also a strong relationship between crude oil prices and Google Trends, indicating that public attention increases when the oil prices show large movements. The increase in public attention is particularly strong when crude oil prices decline. Again, this is illustrated in the post-Global Financial Crisis in 2009, the oil price decline in early 2015 and 2016, and the Coronavirus outbreak in 2020.

In terms of contribution to the literature, Chapter 2 contributes to the literature on exploration and crude oil price in two ways. First, by studying the effect of exploration activity and second, by examining how they respond to the oil price movements. The existing literature uses the common measures for exploration activity, namely rig counts and exploration wells (Dahl and Duggan, 1998; Kemp and Kasim, 2006; Mohn, 2008; Mohn and Osmundsen, 2008; Kellogg, 2011; Ringlund et al., 2008; Anderson et al., 2014; Toews and Naumov, 2015; Skjerpen et al., 2018). Mohn (2008) uses exploration wells as a proxy for exploratory effort; this yields efficiency through expected success rate and average discovery size. As Mohn (2008) suggests, it is essential to identify discovery rate and size separately to understand the sign of the exploratory effort and efficiency to the oil price fluctuation. Thus, this thesis conducts the simulation exercises by applying Monte Carlo to understand the expected time in between two discoveries are made and the discovery size distribution in a mature Norwegian Continental Shelf (NCS). The time between two discoveries (measured in days) results from the exploration activity, depicting the lag between discovery and subsequent discovery. It has the advantage of being easily interpreted by prospective investors or the public, and their investment decision-making processes can gain insight from this measure. Applying a mature petroleum province, where peak exploration and production cycles have been reached and are gradually declining, helps the researcher approach the particular distribution type of the petroleum region.

The simulation utilises NCS discovery size and the time in between two discoveries are made for data from July 1967 to September 2019, to which the Monte Carlo simulation is applied. Monte Carlo simulation employs a large number of random variables to replicate a complex real system (Thomopoulos, 2013). The benefit of applying Monte Carlo is that it takes into account uncertainty in the future discovery size and the time in between two discoveries output. Monte Carlo simulation with 10,000 replication generates the likelihood of the future discovery size and time in between two discoveries to understand exploration activity behaviour through a statistical distribution. The Information Criteria for the Goodness-of-Fit tests result indicate that the most suitable distribution type is lognormal for NCS discovery size, while exponential distribution is most appropriate for the time in between two discoveries. The simulation allows the parameters to vary to capture changes in exploration activity by looking at three scenarios: base case, lower, and higher value of parameters. First, the base case scenario uses the NCS historical data parameters, log mean and standard deviation for discovery size, and the rate of time between discoveries. The simulation repeats this in a second step but choosing lower value parameters, while the final step is to apply values with higher parameters than the base case but drawn from the same distribution.

A toolbox for Monte Carlo simulation is developed in Chapter 2. This can be used to study the distribution of other oil-producing regions. For the NCS case study, the simulation findings indicate that it is more frequent to discover oil in a shorter time between one and subsequent discoveries with a smaller average discovery size. The more exploration activity, the shorter will be the time in between a specific and subsequent discoveries, and the higher the probability density. High exploration activity is represented by more discoveries in a shorter time between two discoveries. Based on the base case scenario, the simulation result shows 90% probability that time between discoveries will be shorter than 316 days. For the same probability percentage, the high value of the parameter (i.e., a high exponential rate) indicates the time in between two discoveries to be shorter than 255 days, while for the low parameter scenario, it is shorter than 339 days. The larger the discovery size, the lower the probability density of finding it. The simulation exercise aligns with the mature petroleum province, where giant oil discoveries are found infrequently and at the beginning of the cycle. These are followed by discoveries that are more frequent but of a smaller size. The base case scenario shows that the 90% probability of discovery size is for less than 93.67 million standard cubic metres of recoverable oil equivalents (Sm^3 o.e.). Looking at the high parameter value, the results indicate that the discovery size is less than

137.01 million Sm^3 o.e. In contrast, the low discovery size scenario results in the discovery size being smaller than 76.11 million Sm^3 o.e. These results are in line with statistics released by Norwegian Ministry of Petroleum and Energy and Norwegian Petroleum Directorate (2022) that the total preliminary estimate of discoveries in 2021 is 85 million Sm^3 o.e. This implies that confidence can be placed in the ability of the Monte Carlo simulation to predict discovery size.

The second part of Chapter 2 analyses how the exploratory efficiency variable reacts to the oil price changes. A structural break is taken into account to capture the shocks in the data series. The later life-cycle (more mature basin) is associated with a smaller discovery size, referring to oil's characteristic as an exhaustible resource. A high oil price stimulates more exploration activity by increasing the number of exploration wells or rig counts. Consequently, this study expects that oil price positively affects exploration activity, reduces the time in between discoveries, and positively affects the discovery size. The empirical estimates conclude that a long-run equilibrium relationship is present between exploratory effort, efficiency, and crude oil price. Incorporating the Auto-Regressive Distributed Lag (ARDL) model with a structural break confirms the presence of the long-run relationship between exploratory effort, efficiency, and crude oil price. Taking into account a break year in 2008, the ARDL empirical results suggest that a 1% increase in oil price is associated with a one-day reduction in the time between discoveries in the long run. A 1% increase in the real oil price is associated with an increase in discovery size by 1% in the long run.

Understanding the response of exploration to the oil price shocks has implications for policy by providing investors and government with insights on how long it takes to make a new oil discovery. It is suggested that policy instruments take this into account when boosting exploration activities. The feedback loop between oil exploration, production, and crude oil price motivates further discussion on the interaction between supply, demand, and price. Understanding oil production response to high and low oil prices in the short and long run also helps with decision making about how much oil to produce, how much to store, and whether to sell now or later. Despite the development of other sources of alternative energy, crude oil exploration and production still play a crucial role in ensuring the security of supply such that oil supply can meet the demand. Recognising and incorporating uncertainty in the global oil market helps predict other pieces of information that are driven by unanticipated events. The empirical results of these studies can be used in sensitivity analysis to help the decision-making process.

Furthermore, the results from the analysis of NCS discovery create a different approach to studying the effect of exploration activity. Chapter 2 provides a reproducible Monte Carlo simulation toolbox to simulate oil discovery size and time in between discoveries that can be applied, through expert judgment, to other regions with or without historical data.

The application of the Chapter 2 analysis is for disaggregated data, in that the discovery data for a mature petroleum province is tailored to the geological characteristics. In contrast, Chapters 3 and 4 apply the time-series model to worldwide data to understand oil market behaviour within a broader global scope. The critical role of the exploration stage motivates further analysis of the oil production stage, which represents oil supply response to oil price shocks. This analysis is presented in Chapter 3, which analyses the relationship between supply, demand, and oil price.

Chapter 3 contributes to the oil prices, supply, and demand literature by presenting asymmetric analysis to understand whether a positive or a negative shock has more significant and dominant effects on the global oil market. It also examines the presence of long-run relationships. The second contribution of Chapter 3 is its analysis of time series properties, most particularly with regard to whether or not the series are stationary and whether there is a structural break in the relationship across the oil prices, suuply, and demand.

It is essential to understand time series properties because failing to specify the appropriate model leads to misspecification in determining the global oil market model. The time series properties are examined by applying the structural break unit root test to accommodate the extreme jump and drop in the series. In addition, a structural break is applied in the relationship across the oil prices, supply, and demand. This chapter applies the specific model of the structural break unit root test proposed by Enders and Lee (2012), which offers a better approach for examining to whether the type of the break is sharp or smooth. Another benefit of applying this test is that the researcher does not need to have *a priori* knowledge of how many breaks occurred in each series. Supremum Wald test estimates the structural break of real oil prices occurred in January 2009 and this is taken into account in the empirical model as it has a strong reason, following Global Financial Crisis. The results confirm mixed order of integration among oil market variables. An ARDL model is carefully applied to allow mixed stationary and non-stationary series. The analysis presents the basic ARDL and the asymmetry ARDL (i.e., the non-linear ARDL or NARDL) models. The basic ARDL model assumes that the shocks among variables are symmetrical. In the asymmetry model, on the other hand, the shocks are decomposed into partial positive and negative shocks. The presence of a long-run relationship is analysed in each model.

The finding indicates a significant relationship between oil supply, demand, and oil price in the long-run equilibrium. Adding a break in January 2009 confirms the strong significance of the long-run equilibrium relationship across those three variables. It emphasises that supply and demand play a major role in the real price of crude oil in the long term, something that has mostly been underestimated in previous studies. The empirical analysis uses data on world oil production, Baltic Dry Index (BDI), and real crude oil price from January 1985 to December 2019. BDI is a bulk dry freight rates-based index, and it is a leading indicator of fluctuation in the global business cycle, indicating global demand. The index covers 31 shipping routes worldwide, using various vessel capacities and specifications.

The long-run relationship between world oil supply, global demand, and crude oil price is present in the basic ARDL model. When they are too high, relative to the equilibrium, they will be adjusted to decrease in the next period so as to return to equilibrium. In the asymmetric ARDL model, the long-run relationship between oil demand, oil supply, and oil price is present and significant. However, positive and negative demand shocks have an effect of similar magnitude on supply and oil price, so the effect is somewhat symmetrical. Oil prices shocks on supply and demand are asymmetrical, as are supply shocks on oil price and demand. The results present a negative relationship between world oil supply and oil price, and a positive relationship between global demand and oil price in the long run. A positive relationship is found between world oil supply and demand. The main findings in the asymmetric analysis are that positive crude oil price shock has contributed more than negative shock to global oil supply and demand in long-run equilibrium. Further, positive oil supply shock has a stronger effect than negative oil supply shock on global demand and oil price, supporting the view of asymmetry. The positive global demand shock is as strong and significant as the negative one, which causes a more symmetrical effect on the global supply and oil price, even though the long-run relationship is still substantial.

Incorporating the ARDL and NARDL models with a break gives a clear understanding of the effects of the break on crude oil prices, supply, and demand and whether positive or negative shocks have stronger interaction with the break. The interaction between the break and crude oil prices and between the break and global demand is significant. The interaction between the break and crude oil production is insignificant in the short run. In the asymmetric model, positive and negative demand shocks are significant on the oil prices, while negative supply shock is found to be stronger in affecting oil prices than the positive supply shock. Both positive and negative supply shocks significantly affect global demand, and none of the breaks is significant to the supply. The supply is less responsive to changes in economics compared to crude oil prices and global demand. World oil production cannot be adjusted quickly, and it may take longer to respond to economic changes.

The long-run relationship and asymmetric effects discussed in Chapter 3 offer some implications. First, though oil production is insufficiently flexible for it to be adjusted in the short term, a mechanism to tailor it in the long term is necessary to overcome high crude oil prices. Maintaining the security of supply is also crucial; in this case, it is necessary to boost world crude oil production to ensure that supply meets the strong global demand. These implications are also linked to Chapter 2 in that exploratory effort and efficiency are also required to boost additional reserve growth. Here, finding new oil fields and having a shorter time between discoveries plays an important role. Second, as demand shocks have symmetric effects, oil prices are expected to fluctuate whenever strong or sluggish global demand occurs. The oil exploration and production companies and governments can use the high demand signal to speed up exploration and production activities, and they can thus benefit from the high oil prices. In contrast, policymakers need to consider an effective inventory mechanism to overcome low demand and low oil prices. Regular reviews of the estimated and actual crude oil storage utilisation rate are necessary to minimise the risk of oversupply.

The findings from Chapter 3, which emphasise the importance of demand shock and positive supply shock on oil price fluctuation, support the earlier work by Baumeister and Kilian (2016*b*). Their study argues that the cause of the oil price drop between June and December 2014 was 55% predictable due to positive oil production and negative global demand shocks. They also conclude that the remaining 45% was attributed to unpredictable shocks related to oil price expectation and unexpected slow global demand. This unpredictable shock is interpreted in terms of market concerns associated with uncertainty, which is discussed in Chapter 4.

The considerable attention given by the literature to measuring uncertainty has tended to adopt a macroeconomic perspective, which has resulted in the macroeconomy and policy uncertainty measure. The traditionally applied uncertainty index is the volatility index, a commonly used index based on the short-term volatility expectation as priced by the United States Oil Funds, such as OVX for crude oil price and VIX for stock market volatility. The earlier literature applied conditional forecast error based on oil price change to propose oil price uncertainty, as seen in work done by Lee et al. (1995) and Elder and Serletis (2010). The remaining studies mostly apply forecast-based or newspaper coverage-based analysis for macroeconomy, policy, or financial uncertainty. Jurado et al. (2015) use common macroeconomic indicators and apply conditional volatility of future values to construct a macroeconomic and financial uncertainty index. The recent literature measures economic policy uncertainty through a newspaper-based approach. Baker et al. (2016) construct the economic policy uncertainty index by quantifying the frequency of terms related to economic, policy, and uncertainty in the newspapers of various countries. The US economic policy uncertainty index finds that policy uncertainty is associated with larger stock market volatility. Davis (2016) extends the work of Baker et al. (2016) by incorporating the weighted-average GDP of 21 countries to measure national economic policy uncertainty in a global index.

The newspaper coverage-based uncertainty index concept can be adopted for an internet search-based index. It derives from the notion that individuals search for what they are interested in by typing search terms (i.e., keywords) into a website browser's search box. Keywords with high relative frequency reflect higher public interest in these terms. Most studies follow prior non-internet-based literature in their construction of internet search-based indices for the macroeconomy, policy, or general uncertainty. Castelnuovo and Tran (2017) provide a comprehensive procedure for constructing a Google search-based macroeconomy uncertainty index. A few studies have applied a Google search-based index for oil price uncertainty, such as Qadan and Nama (2018); Li et al. (2019); Guo and Ji (2013); Ji and Guo (2015). However, Qadan and Nama (2018) and Li et al. (2019) only apply a basic oil price index to measure investor sentiment. The studies of Guo and Ji (2013) and Ji and Guo (2015) are comprehensive in terms of the oil market literature. However, some search terms in Guo and Ji (2013) are not applied correctly when it comes to approaching oil demand uncertainty; for instance, the search term 'oil production' measures oil demand and 'gas price' measures oil price uncertainty. Likewise, Ji and Guo (2015) use the search terms 'Libyan war' and 'OPEC conferences' to approach expectations on the oil price, even though these search terms are directly associated with oil supply rather than with market expectations of the oil price.

Motivated by the absence in prior literature of an index for global oil market uncertainty, Chapter 4 contributes to the literature in several ways, the first of which is to use Google Trends to capture public interest in specific aspects of oil prices via five newly-constructed indices: oil price, oil supply, oil demand, oil investment, and oil market-specific, with only the oil price index being a basic model. The remaining indices are more nuanced, being constructed by considering important keywords related to oil supply, oil demand, and oil investment. Second, Chapter 4 analyses the uncertainty impact on oil market-specific variables, most notably crude oil price and oil exploration; this perspective is relatively unknown in the existing studies, which are mostly interested with uncertainty's effect on macroeconomic variables. In this study, North America's rig counts are analysed separately from worldwide rig counts. North America plays a major role in the non-conventional drilling for exploration and production, and its volume is significant compared to the rest of the world.

Third, Chapter 4 carefully compares the current macroeconomic, financial, and economic policy indices and the newly proposed uncertainty measures to understand the performance of the index in predicting global oil market variables. Fourth, the newly constructed index gives an overview of worldwide oil market uncertainty; in this, it differs from a country-specific index that is restricted to the application of certain countries.

The application of Google Trends measures public interest. When there is uncertainty, public interest in oil prices is high. The newly proposed index relies on psychological behaviour in the digitised era, where information is generally sought from the internet. As the most widely used web browser, this chapter chooses Google search. Google Trends is a web-based free service provided by Google that indicates the relative frequency of search queries to identify the most popular (i.e., trending) search terms. In addition to being free, open to the public, and easy to navigate, Google Trends offer flexibility. For instance, the researcher can compare up to five search terms in each round, can filter the data according to region, and can extract the frequency (from daily up to yearly). Worldwide data is chosen as the focus of the study, from which a global benchmark uncertainty measure is proposed. The measure improves understanding of the causality relationship between exploration activity and oil price.

The first section of Chapter 4 describes the important contributions of the newly proposed Google Trends-based Uncertainty (GTU) index, comparing its correlation with the established uncertainty indices, such as the oil volatility index and indices related to the macroeconomy, global economic policy, geopolitical risk, and world uncertainty. This study applies monthly data from January 2004 to April 2020 to construct five proposed GTU indices; these go from basic to more refined. The five indices are constructed by carefully selecting search terms to measure public interest in the global oil market, basic GTU oil price, GTU oil supply, GTU oil demand, GTU oil investment, and

GTU oil market-specific. The final measure is the most complex, aggregating the first four components.

The contribution of uncertainty impact on oil market-specific variables is analysed in the second part of Chapter 4. The existing studies are mostly interested in the uncertainty effect on macroeconomic indicators. Jurado et al. (2015) conclude that uncertainty shocks cause a large decline in real activity, such as production, hours worked, and employment. Baker et al. (2016) also find that policy uncertainty shock reduces investment, output, and employment in policy-sensitive sectors. A vector autoregression model is applied because of the interest in examining the dynamic relationship between uncertainty shock and oil market-specific variables in the short term. Public interest will become particularly intense when there is uncertainty due to unexpected circumstances, and this type of uncertainty shock tends to occur in the immediate and short term. The estimation expects GTU indices to have consistent results compared to the benchmark uncertainty impulse response for oil price and exploration. Eight separate models allow the researcher to compare how the different uncertainty measures affect crude oil price and exploration. Each model consists of certain essential variables: the number of rig counts, crude oil price, and one uncertainty measure. There are three existing benchmark indices (oil volatility, macroeconomy, and global economic policy uncertainty) and five newly proposed ones.

The contribution of Chapter 4 in carefully comparing the existing and the newly proposed uncertainty measures are applied by providing correlation measures, time series plots, and impulse response functions. First, the newly proposed GTU indices positively correlate with the benchmark indices, which are the oil volatility and macroeconomy uncertainty indices. GTU indices can capture unexpected events associated with crude oil price spikes and drops during, for instance, the global financial crisis 2008–09, the oil price drop in 2014–16, and Coronavirus in 2020. Second, GTU indices have signs that are consistent with the benchmark indices in the regression relationship with the oil price and rig counts. Thus, uncertainty has significant negative effects on oil price fluctuation and rig counts; oil price shocks negatively affect uncertainty but this is only significant for a few uncertainty indices; and oil price shocks positively and significantly affect the rig counts of the world and North America. The rig counts positively and significantly affect oil price, particularly in the world model, but they do not significantly affect uncertainty.

The main contribution of Chapter 4 is the construction of the GTU index capturing public interest in the global oil market, which is intensified when there is uncertainty due to remarkable events in oil price episodes. The internet search-based engine as a transmission channel links market concerns in supply, demand, investment, and the oil market. This newly constructed index can explain the variation in the fluctuation in oil price and exploration activity, which emphasises the role of uncertainty on oil price shock as illustrated in Figure 1.1.

The concluding chapter of the thesis summarises the main findings and the lessons learnt from the analysis and discussion in the three empirical chapters of oil price shocks, exploration, production, and uncertainty. It also outlines how this study benefits researchers, policymakers, investors, and business sectors, particularly those related to the oil business's upstream. The insights into the economic policy implications and further research directions consequent to this research are presented in this chapter.

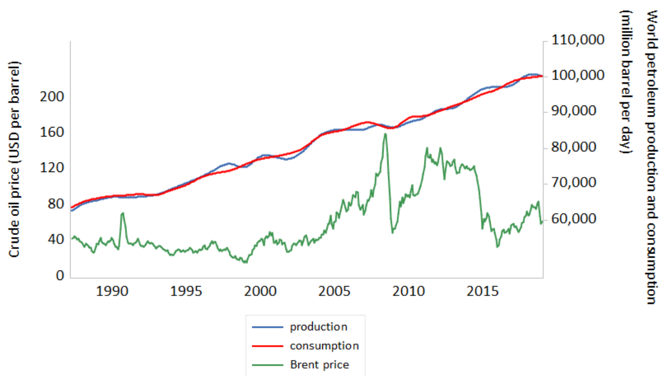
Chapter 2

Exploration activity and the crude oil price: evidence from the mature petroleum province

2.1 Introduction

Since the early 1970s, there have been at least eight extreme oil price episodes: 1979–81, 1985–86, 1990–91, 1998, 2001, 2008–09, 2014–16, and 2020–22. It is therefore unsurprising that research is interested in uncovering the main drivers of oil price fluctuations. Hamilton (2011) explains that the key factors contributing to previous oil price fluctuation episodes are related to supply disruption arising from geopolitical conflicts in the Middle East. For instance, oil price peaks in 1980 and 1990 are associated with the Iranian revolution and the first Persian Gulf War, respectively. In addition, growth in petroleum demand combined with declining production is another contributing factor, as seen in the extreme oil episode in 2007–08, when mature petroleum provinces such as the UK, Norway, and the Gulf of Mexico were in decline. An associated major event in 1997–98 was the Asian financial crisis, which caused a reduction in oil demand, which has been presumed to be the cause of the crude oil price drop at that time (Baumeister and Kilian, 2016*a*). The decline in oil price between 2014 and 2016 can be explained by the expected positive shock in oil supply (33% of the total decline) and negative shock in demand (22% of the total decline), while the remainder is due to unexpected oil market expectation (Baumeister and Kilian, 2016*b*).

The interaction of petroleum demand and supply as an explanation of extreme oil price fluctuations can be visually explained (see Figure 2.1). The global petroleum supply and demand data reflect these findings; the lines representing global consumption and production exhibit an upward trend. They are often very close to each other and most of the time they intersect. However, production rises between 1997 and 1998 are associated with lower consumption, and crude oil prices dropped. Other gaps occur in 2006–07 and 2009–10, when consumption is more prominent than production. In these periods, oil prices exhibit an extreme increase. Later, between 2014–16, production quantities outstrip consumption and oil price experiences another significant decline during this period.

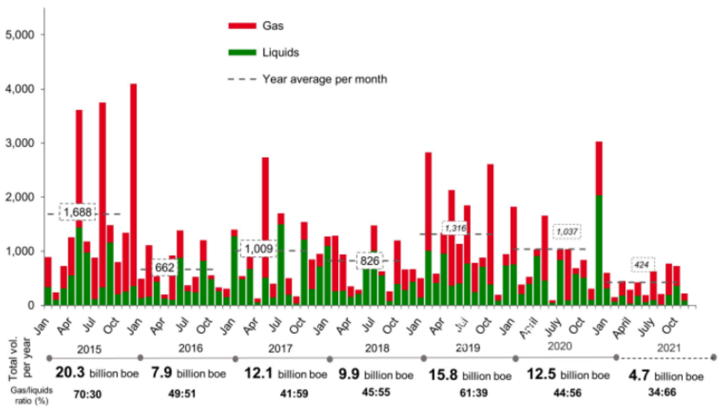


Source: EIA (2021), amended from Gronwald (2016)

Figure 2.1: Global petroleum production, consumption, and crude oil price

So much for the well-established effect of oil supply on oil prices. This chapter, however, is concerned with the opposite effect – that of oil price changes on oil supply. This research's motivation derives from recent micro-level research that indicates that it is oil exploration rather than crude oil production that responds to oil price increases (Anderson et al., 2018; Mauritzen, 2017). Among the extant work that uses drilling activity as the proxy for exploratory effort, Toews and Naumov (2015) key empirical findings are that 'a 10% increase in the oil price significantly increases global drilling activity by 4%'. Mohn and Osmundsen (2008) find a significant long-term relationship between drilling effort and crude oil price in the Norwegian Continental Shelf (NCS), and they also note that the long-term elasticities of oil price on exploration are modest in a highly regulated petroleum province with high taxation. Anderson et al. (2018) proposes a modified Hotelling resource extraction model, noting that the resource owner does not decide when to produce oil, but rather when to drill; it is geological constraints that determine production from existing wells. While Anderson et al. (2018) use data from Texas, Mauritzen's (2017) empirical study analyses NCS production data and argues that production in the existing field is not affected by high oil price in the short term but rather by a lag, estimated to be between two and four years. When it comes to short-term hedging against oil price fluctuation, oil producers prefer to use financial instruments or storage.

There is a feedback loop between crude oil price turbulence, oil supply, and demand. Supply and demand determine today's price, but there is also a link between today's and tomorrow's supply because of the exploration channel. Thus, current exploration activity is an essential determinant of future supply, which makes a critical contribution to oil price movements. Depending on which direction the global economy takes, and how future crude oil demand develops, there may be further extreme oil price periods in the future. Industry observers such as McKinsey & Company (2021) report that new crude oil production of 38 million barrels per day, or 23 million barrels per day under the energy transition scenario, is required to meet demand in 2040. Rystad Energy, a Norwegian energy consultancy, reports that the global oil and gas



Source: Rystad Energy ECube, UCube, research and analysis (December 2021)

Figure 2.2: Global discoveries for 2021 on course to lowest in decades

discoveries were equivalent to 4.7 billion barrels of oil in 2021, the lowest level since 1946 (see Figure 2.2). The low size of discoveries has started to evince concerns about a supply gap in the crude oil market. Despite these trends, the relationship between exploration and oil price is insufficiently recognised in the academic literature, although there are signs of improvement.

Various methods of estimating oil discovery and production have been proposed, including geological engineering, mathematical geoscience-based, or economics-based forecasting. Hubbert (1962) portrays the oil discovery and production profile as a bell-shaped curve with a single peak during the cycle. Later, a creaming method was developed; this is based on proportional sampling between the probabilities of successfully making a discovery and the size of the discovery; this method become well-known for forecasting resource size. Kaufman et al. (1975); Smith (1980); Lillestøl and Sinding-Larsen (2017) all apply the creaming model. This model finds that the common distribution profile of a mature petroleum province is a lognormal distribution, in which the giant discovery size is located in the earlier period. Smith and Paddock (1984) propose a discovery model in which the decline in expected discovery size follows exponential decay. MacDonald et al. (1994) propose a nonlinear polynomial regression to forecast the discoveries volume in Canada's Western

Sedimentary Basin. These methods are based on the petroleum fields' geological and statistical properties, so they fail to explain the exploration response to oil price shock. When the research objective is to examine how oil price movements affect oil discovery, further econometric estimates are required to link these with oil price.

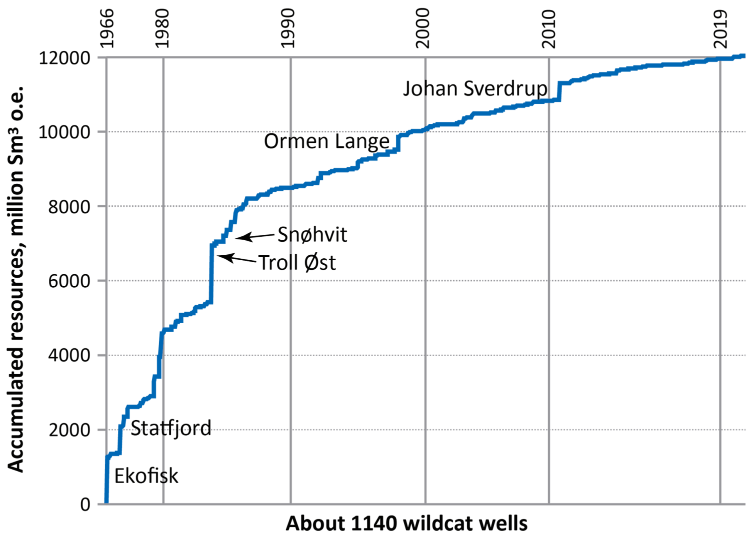
Recent studies in oil discovery forecasting contribute to the literature by applying the time-series econometrics model. The UK Continental Shelf (UKCS) and NCS are the North Sea's two most productive oil reserves. Kemp and Kasim (2003) integrate physical and economic variables for the UKCS application. Their work emphasises the importance of the feedback effects of exploration and appraisal, development, production, reserves, price, and investment. Kemp and Kasim (2006) extend previous work by incorporating exploration cost and efficiency. Mohn (2008); Mohn and Osmundsen (2008) make contributions to the analysis of exploration activity in the NCS in their work on exploratory effort and efficiency for highly regulated oil provinces. Their studies define exploratory effort as drilling activity represented by the number of exploration (i.e., wildcat and appraisal) wells. Their comprehensive works take into account licensing policy and technological changes, as well as effort and efficiency.

Two research questions are addressed in this chapter. First, how does exploration activity behaviour over time? Second, how does crude oil price fluctuation affect exploration? Then, this chapter answers the research questions and advances research in this area via two steps: The first numerical step looks at how exploratory effort leads to exploratory efficiency. The exploratory effort mainly affects the frequency of discoveries, which is, in turn, determined by factors such as the oil price and tax regimes. This chapter studies how the frequency of discoveries is translated into exploratory efficiency, adopting the definition of exploratory efficiency as the yield of exploratory effort, which is discussed in the earlier studies by Iledare and Pulsipher (1999); Kemp and

Kasim (2003); Mohn (2008), and Mohn and Osmundsen (2008). The second step is empirical estimation and it analyses the relationship between oil prices, exploratory effort, and efficiency. The second step empirically analyses the relationship of the geological-based variables applied in the first step and takes into account the real oil prices as an economic variable. Most of the earlier literature defines it as the proxy of exploratory effort by how many oil rig counts or oil exploration wells are drilled. The exploratory effort is affected not only by the petroleum region's geographical condition but also by economic variables, such as crude oil price (Anderson et al., 2018; Toews and Naumov, 2015; Mauritzen, 2017). Thus, it motivates the second step to carry out an empirical estimation for waiting time, discovery size, exploration well counts, and crude oil price. The third and final step would be to analyse how discoveries translate into future production, but this is a matter for further research. However, the chapter takes into account the very long lags between the discovery of an oil field and the beginning of production from that field; in the case of the NCS, this lag is longer than three years and can take up to fifteen years (Norwegian Ministry of Petroleum and Energy and Norwegian Petroleum Directorate, 2022).

The first part of this study captures the exploratory efficiency as the result of the exploratory effort by modelling the time between successful oil discoveries and the discovery size. The time between successful oil discovery is interesting from an economic perspective as it indicates the result of the exploratory effort, which determines future supply, as well as an indicator of the result of investment in oil exploration. The shorter time between discovery accelerates oil production, and as a consequence, it increases the supply side. This study expects that high oil price reduces the time between oil discovery (waiting time) in the economic model. EIA (2019) reports that low oil price leads to the delay of the time to complete a well for economic reasons. This thesis applies the time between oil discovery to represent exploratory efficiency, contributing to the existing literature, which mainly uses Yield per Effort (YPE), average success rate, and average discovery size. The earlier research in the oil exploration model applies Logistic Curve Fit by Hubbert (1962) and Exponential Decline Curve Fit by Hubbert (1967). Hubbert (1967) uses cumulative drilling to construct YPE measuring the quantity of oil discovered per foot of exploratory wells drilled. It uses the Hubbert Unit, measured in 108 feet of exploratory drilling, instead of the time unit for the Exponential Decline Curve Fit. The Hubbert Unit does not give information on the year to a certain discovery size. Therefore, it motivates this study to provide information about the discovery year associated with the discovery size for a more straightforward interpretation.

The current literature focuses on cumulative exploration and production to predict the remaining reserves in a particular region (Smith, 1980; Reynolds, 2002; Fiévet et al., 2015; Lillestøl and Sinding-Larsen, 2018). In comparison, this study makes innovative use of the Monte-Carlo simulation to answer the first research questions by obtaining the predicted waiting time between discoveries and size for a sequence of discoveries in the mature petroleum provinces. The frequencies of waiting time between discoveries and discovery size generated form the suitable distribution and help to understand the behaviour of the exploratory efficiency over time. The approach applied in this study is generally flexible, as different distributions for discovery size can be used to model other oil-producing regions. The datasets are taken from the 141 sequence discoveries in the NCS from 1967 to 2019. NCS data was chosen because Norway has initiated wildcat exploration drilling since 1966. Most mature petroleum provinces have reached the peak of their exploration and production cycles and are currently gradually declining. Thus, the researcher can more readily approach the particular distribution type of the petroleum region. Although it is well known that the size of the discovery follows a lognormal distribution, this study uses a statistical test to choose the proper distribution, this being crucial



Source: Norwegian Ministry of Petroleum and Energy and Norwegian Petroleum Directorate (2022)

Figure 2.3: Accumulated resources on the Norwegian continental shelf

to the goodness–of–fit of simulating the discovery profile. The distribution also takes into account that the larger size of promising fields are discovered in the early basin life-cycle, as illustrated by the historical trend of NCS discoveries where the large discoveries were mainly made at the beginning of the exploration phase in 1969 (Ekofisk), 1974 (Statfjord), 1979, and 1983 (Troll). Note that there are a few exceptions to this, such as Ormen Lange in 1997 and Johan Sverdrup in 2010, as illustrated in Figure 2.3.

The Monte-Carlo simulation result provides a clear link between exploratory effort and efficiency parameters. Initially, a longer time is required to discover a larger size field in an oil province; then, discovery size gets smaller albeit the discoveries are more frequent and there are shorter waiting times between them, reflecting high exploratory effort. Consequently, it takes a shorter time to reach the last discovery. The NCS simulation result shows that lognormal and Weibull distributions have good fit with the discovery size of NCS. Meanwhile, exponential and gamma distributions have a good fit for simulating the waiting time between a specific discovery and the one that is subsequent to it.

The second part of this chapter conducts an empirical estimation to analyse the interaction between exploratory effort, efficiency, and oil price. It contributes to the literature by splitting the waiting time between discovery and discovery size to clearly understand the effect of oil price fluctuations on exploratory effort and efficiency, as suggested by Mohn (2008). The exploratory effort variable is taken into account to ensure consistency with the existing literature. The waiting time between discoveries and discovery size is regressed on the exploration wells and crude oil price by applying an Autoregressive Distributed Lags (ARDL) model. The long-run relationship and short-run dynamics are distinguished in the empirical model. The empirical estimates show the importance of distinguishing between waiting time and discovery size.

A structural break is taken into account to capture the negative shock in the sample period. A long-run relationship is found between exploratory effort, efficiency, and oil price in both basic ARDL and ARDL with break models. A break in 1981 is found for the waiting time between discoveries and a break in 1999 for discovery size. A break year in 2008 is also estimated in the model as the series contain extreme shocks by crude oil price during Global Financial Crisis. The model with a structural break confirms the presence of the long-run relationship between exploratory effort, efficiency, and crude oil price. The empirical results are consistent with the simulation exercise regarding the relationship between exploratory effort and efficiency. Crude oil price and the number of exploration wells relate negatively to the waiting time between discoveries in the long run. When the oil price is high, it stimulates exploration activity, leading to more exploration wells and a higher frequency of discoveries. The more frequent the discoveries are, the shorter the waiting time between them. Crude oil price relates positively to discovery size, whereas exploration well number relates negatively to it. This result is consistent with Mohn (2008), who posits that the positive relationship between oil price and discovery size is due to the incentive to discover the bigger new resources that are mainly found in a frontier petroleum province. As exploration wells increase in number and are more frequently discovered, the discovery size gets smaller. In the short term, crude oil price does not significantly affect exploratory efficiency, whereas exploratory effort does.

The remainder of this chapter is organised as follows: Section 2.2 provides some background and reviews the existing literature. Sections 2.3 and 2.4 present the data and methodology, respectively. Section 2.5 illustrates the steps to select the underlying distribution type of waiting time between discoveries and discovery size. Section 2.6 reviews the simulation and empirical models for exploratory effort, efficiency, and crude oil price, and section 2.7 concludes the analysis.

2.2 The literature on exploratory effort and efficiency

Exploration activity is an important stage in a future oil supply as the current exploration determines the oil production rate. Besides, exploration activity also affects the oil reserves as oil is an exhaustible resource. The importance of exploration activity motivates various approaches in previous research to estimate the remaining or undiscovered reserves. Geological and economics-based are two strands in the well-established modelling framework in oil exploration. This section describes what existing literature mainly focuses on and the methodology applied to estimate the reserves.

Many established works of literature have attempted to model exploratory efforts by applying various proxies. Several proxies to measure exploratory efforts include the number of wells drilled, exploratory footage drilling, time trend (Arps et al., 1971), discovery rate, cumulative discoveries, resource price, exploration expenditure, finding, and extraction costs (Power and Jewkes, 1992). Arps et al. (1971) applies cumulative footage drilling as the independent variable to predict the ultimately recoverable resources. Osmundsen et al. (2010) use the average drilled meters per day as the proxy of drilling productivity for Norwegian oil activities and conclude that oil price increase causes the decline in drilling productivity. Swierzbinski (2013) defines two factors; cost of extractions and resource prices are essential to measuring resource scarcity. Contrary to the benchmark Hotelling's model that the cheapest cost is extracted at the early phase and cost rises over time for the exhaustible resources, Swierzbinski (2013) argues that technological advance leads to decreased extraction cost over time.

The most applied methodology to forecast discovery and production is the curve fitting technique, including the growth curve. The initial work in geological-based modelling is by Hubbert (1962). Hubbert (1962) initially proposed the US ultimate recoverable resources and production rate projection through the growth curve. Hubbert's model is well-known as a symmetric bell-shaped production curve. Some critics of Hubbert's model are that the reserve growth has a different magnitude on its rate than the reserve decline rate, leading to the asymmetric discovery and production curves rather than the symmetric one (Sorrell et al., 2010). Arps et al. (1971) emphasises two fundamental thoughts in the exploration concepts: decline in drilling efficiency and the growth in future discovered reserves. The higher the exploratory efforts, the higher the exploration results in a steep increase at the beginning and end with a more gradual rise up to the ultimate recoverable resource with some economic limit as the constraint. An economic limit becomes the threshold between the proved reserves and the accumulation of undiscovered resources.

The recent research applies the creaming curve technique, including Lillestøl and Sinding-Larsen (2018), that uses a creaming curve with maximum likelihood estimation to obtain a reasonable prediction within the mature part of the Norwegian Continental Shelf. Hallack and Szklo (2019) apply a creaming curve variant to assess exploratory potential in Brazil. They argue that a creaming curve is suitable to model the petroleum province where the field data are limited. In their model, the creaming curve variant considers a recovery factor to estimate a discovery size by field. On the contrary of curve fitting applicability in forecasting, Sorrell and Speirs (2014) argue that curve fitting techniques tend to underestimate the ultimately recoverable resources. The curve fitting technique neglects factors other than geological aspects.

Another commonly applied methodology to estimate the size of remaining reserves than curve fitting is a probabilistic distribution. The existing research classifies the oil field size based on whether the resources are technically and economically viable to recover. However, there is no consensus regarding the giant and dwarf fields thresholds, and every author justifies their choice. Some research classifies 50Gb (Höök and Aleklett, 2008); 50Mb (Fiévet et al., 2015) for the giant fields. Barouch and Kaufman (1976) proposes a postulate for sampling without replacement from the hydrocarbon pools for Canada that makes a future discovery conditional on the previous discovery size. Arps and Roberts (1958) also propose that the frequency-density distribution can estimate the expected future drilling, and the drilling process is random. Adapted from Kaufman et al. (1975) and Barouch and Kaufman (1976), Eckbo et al. (1978) and Smith (1980) extend the work for the application in the North Sea. Particularly, Smith (1980) apply a probabilistic distribution to estimate remaining reserves in the Norwegian petroleum province. Lee and Wang (1983) use Monte Carlo to simulate the probability distributions of random geological variables in the Canadian sediment basin. Fiévet et al. (2015) extend the work of Smith (1980) to model the discovery rate for the discovery in Norway and the UK by Poison distribution process and fit to the logistic curve to obtain the respective time distribution. They follow the two fundamental notions that the discovery rate depends on the size of the oil field, and the resource is finite that discovery rate leads to a smaller size and tends to zero.

Some authors propose the exponential decline of the discovery rate and agree that the distribution type of the discovery size mainly follows a lognormal distribution. Meanwhile, some follow other distributions such as gamma and Pareto distribution.Kaufman et al. (1975) argue that a lognormal distribution is suitable for describing the size of the petroleum reservoirs in Alberta, Canada. Beall (1976) also assumes that the lognormal is reasonable to approximate the petroleum reservoirs in the North Sea. Smith and Paddock (1984) model the discovery size as exponential decay. They assume that the exponential decay rate is based on cumulative historical discoveries, and the expected size of the last discovery would be the minimum economic discovery size. Bohling and Davis (1993) presents oil discovery sequence simulation following lognormal, gamma, and Pareto distribution and concludes that the model with either distribution tends to simulate more large fields than the actual discovery.

The literature investigates the model of the discovery and exploratory efforts in disaggregated data, well or field-based observations. The decision to apply disaggregate data is based on the fact that the geological condition varies among regions. Classifying the study based on the maturity area of the resources is also common. Most of the well-established research considers mature petroleum provinces such as the Norwegian Continental Shelf, the UK Continental Shelf, the US and Canadian basins, the Gulf of Mexico and other developing countries such as Brazil and India. Power and Jewkes (1992) propose a discovery efficiency parameter in applying the Western Canada sedimentary basin. They argue that an improvement in the exploration efficiency can reduce the exploratory effort but not maximise the total discovery when executed in the early phase of the exploration. MacDonald et al. (1994) estimate the cumulative discovery volume in Canada's Western Sedimentary Basin by applying a linear regression of third-degree polynomial. The cumulative discovery volume is forecasted based on the well counts. Höök and Aleklett (2008) investigate the decline rate of Norwegian oil production and found that annually giant fields decline by 13%, and the smaller fields decline rate is faster by 40% annually. Rao (2000) conceptualises an integrated optimisation model of exploration and extraction for the application in India. The research combines the optimal rate of extraction to estimate the discovery rate that results in the simulated production-to-reserve ratio for the policy decision-making process.

The literature then has developed and considered the economic and other factors, such as technological progress and politics, into the model. Ramsey (1980) emphasises the crucial issue of the optimisation exploration process rather than the optimal depletion of the exhaustible resources. Its contribution to the economic literature is by investigating the exploration firms' behaviour and arguing that large firms tend to explore new discovery areas and then proceed by smaller firms until no exploration as the field is decayed. Some extensions of Hubbert's model that incorporate economic factors are Kaufmann (1991), and Cleveland and Kaufmann (1991). Cleveland and Kaufmann (1991) modify Hubbert's original projection on the US ultimate oil recovery and rate of production by accounting for a non-random drilling in Yield per Effort curve and incorporating political and economic factors. Pesaran and Samiei (1995) evaluate the performance of the original Hubberts' model and its extensions by Kaufmann (1991) and Cleveland and Kaufmann (1991) from an econometric point of view.

Reynolds (2002) extends Hubbert's oil discovery and production model to include technology and regulation changes. The model applies a non-time-series cumulative discovery quadratic Hubbert curve and structural shift variables. Forbes and Zampelli (2002) propose empirical estimation of the technological changes on the exploratory success rate. Managi et al. (2006) argue that exploratory effort and technological changes significantly affect the discovery at the field level. They apply the average drilling distance per exploratory well and the number of exploratory and development wells as the exploratory effort proxies in the application of Gulf Mexico.

As the economic factors are incorporated into the exploration model, the study of the relationship between oil market variables, such as crude oil price, exploration cost, and other financial instruments, has developed. Through the empirical analysis, Pesaran (1990) argues that oil price strongly affects UKCS exploration and production. Dahl and Duggan (1998): Oil prices affect geophysical and drilling activities. They approach oil exploration through wildcat wells drilled, the success rate percentage of commercially viable wells, and the average oil reserve per successful well. Farzin (2001) argues that future oil price significantly affects the additions to proven reserves from existing fields in the US. A 10% increase in real oil price leads to a 1% increase in proven reserves. Forbes and Zampelli (2002) also conclude that the oil price has a significant positive effect on the US onshore exploratory success rate. Applying a VAR framework to the US data, Ewing (2017) classifies the discoveries based on its sources; extensions to reserves, new field discoveries to reserves, and new reserves discovered in the oil field to reserves to the shocks of crude oil price and 10-year Treasury bond. Ewing (2017) argues that the extensions-to-proved reserves respond more substantially to the oil price fluctuation than the new discoveries. Meanwhile, the new field discoveries and new reservoirs discovered in oil fields respond positively to the interest rate changes.

Kemp and Kasim (2003) design an econometric model incorporating exploration and production activities, reserves, price, cost, and taxation to apply mature petroleum province of the UK continental shelf (UKCS). Having extended their previous work, Kemp and Kasim (2006) adds the effects of technology and exploration efficiency on discovery. They argue that the cumulative number of exploration and appraisal wells, cumulative discoveries, exploration failure rate, and time trend is significant in describing the UKCS exploration efficiency behaviour. Mohn and Osmundsen (2008) found that a 1% rise in oil price increases Norwegian exploration well by 0.20% in the short run. In the long run, they found that an increase in oil prices from the previous year also leads to the rise of a Norwegian exploration well by 0.21%. Toews and Naumov (2015) applied a structural VAR to estimate the oil price shocks to the drilling activity. They argued that a 10% increase in oil price shocks leads to increased UK exploration well and drilling costs by 4% and 3%, respectively. Skjerpen et al. (2018) argue that oil price and rig rates have a positive relationship, and the oil price has to stay high for some time to affect the rig rates considerably.

This study differs from the previous literature and contributes to the literature. The contribution of this study provides a new alternative proxy for the exploratory effort, the waiting time simulation between discovery. The current establishes literature applies maximum likelihood probability distribution and the exponential decay of historical means of discovery rate that cannot obtain the time distribution. The scope of this research is limited to the exploratory effort of the new discovery fields and does not include the growth of the existing fields.

This study also provides a tool in R-software to simulate the waiting time in a particular petroleum province. The benefit of applying this tool is that it does not require extensive numerical programming. Applying disaggregate Norwegian Continental Shelf (NCS) field-based discovery data, the parameter of the declining rate can be estimated. This study runs the Mean Absolute Percent Error (MAPE) and bias tests to evaluate the accuracy of the simulated values. The other benefit is that when the historical data is unavailable, or there are not enough observations, the predicted parameters from the experts can be inputted into the tools to obtain a simulated waiting time.

The second contribution is that this study analyses the relationship between the exploratory effort applying the waiting time between discoveries and the global oil market variables, remarkably crude oil price. Hence, the linkage between geological and economic factors is not missing. Furthermore, the consistency and reliability of this new proxy can be identified by comparing the results with the well-established exploratory effort variables, such as the number of the exploration well and rig counts.

2.3 Oil discovery data and descriptive statistics

This section is organised into two parts. The first part discusses the data employed in the simulation of Norwegian Continental Shelf (NCS) oil discovery sequences. The second part examines the data used in the empirical analysis of the relationship between exploratory effort, efficiency, and crude oil prices. The main focus of this study is to determine the time between one discovery and the next (i.e., waiting time), and then to create a future discovery simulation profile. This study provides a simulation toolbox for more straightforward flow for estimating a future discovery. Appendix A.1 shows the code of the discovery size and waiting time simulation.

The waiting time obtained from the first part of the section reflects the exploratory efficiency. The next subsection provides the data for the regression of the waiting time on the number of exploration wells, and crude oil price. The average discovery size is included in the second regression to distinguish the effect from time between one and subsequent discoveries. The motivation for the empirical analysis is mainly to understand the consequences of oil price fluctuation on exploratory efficiency, as the result of exploratory efforts.

2.3.1 Norwegian Continental Shelf (NCS) discovery sequences

This subsection describes the NCS discovery sequence as the basis of the simulation process. In this study, the term ‘discovery’ is based on the assumption that successful discovery of a new field will lead to production. Thus, oil production growth from an existing field and the findings from fields that are later abandoned are not considered. The order of discovery, the discovery date, and the size of the field discoveries in million Standard cubic metres of oil equivalents (mill Sm^3) are the three essential components for modelling the sequence of the discovery simulation.

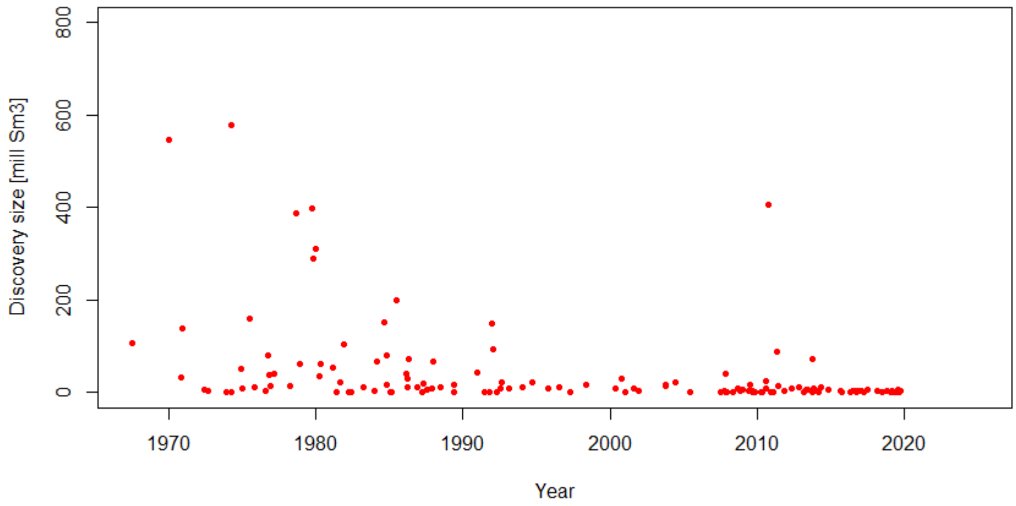


Figure 2.4: Norwegian Continental Shelf oil discoveries 1967-2019

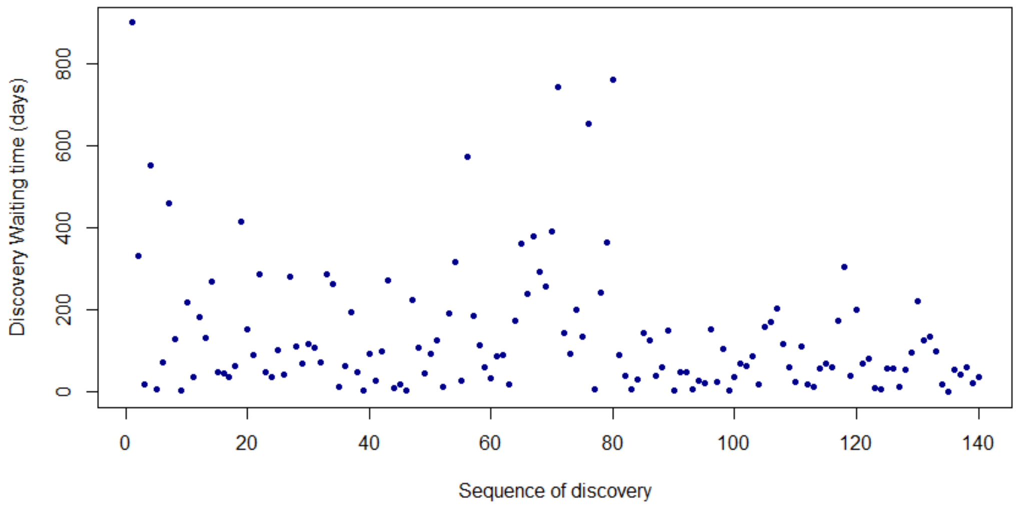


Figure 2.5: Norwegian Continental Shelf oil discovery waiting time

The simulation takes into account the historical data of 141 NCS discovery sequences from July 1967 to September 2019. The discovery dates and sizes are obtained from the Norwegian Petroleum Directorate (NPD) database.[1]

[1]https://factpages.npd.no/en/discovery/TableView/Overview

Figures 2.4 and 2.5 illustrate the time series plot of Norwegian oil discovery sizes and waiting time. The waiting time is the time difference between one discovery and the next, and thus it results from exploratory effort. Based on the probability distribution of NCS discovery sequences, this study simulates the discovery size and waiting time between discoveries. At the beginning of the observations, there are more large discoveries than there are during the rest of the period. Figure 2.5 shows that it takes longer to make a discovery in the early period. However, small size discoveries are made more frequently found than large ones.

Table 2.1: Descriptive statistics of NCS discovery data

Summary statistics	Size (million Sm^3)	Waiting time (days)	Dates
Min	0.02	0	09-07-1967
Max	578.69	899	27-09-2019
Mean	41.08	136.24	20-08-1997
1st Quartile	2.29	37.75	25-12-1983
Median	7.90	88.00	03-05-1998
3rd Quartile	24.10	176.25	27-10-2011
Estimated sd	95.8753	157.7070	
Estimated skewness	3.7873	2.3760	
Estimated kurtosis	18.2672	9.7494	

Table 2.1 contains the descriptive statistics of discovery size and waiting time. The average successful discovery size for the Norwegian field between 1967 and 2019 is 41.08 million Sm^3 (equivalent to 258 million barrels), with an average waiting time between one and the subsequent discovery being 136 days. The standard size for a giant discovery is 500 million barrels Nehring (1978); Zou (2013), although Fiévet et al. (2015) uses the assumption of a giant field size of 50 million barrels. The biggest discovery (namely, Statfjord discovered in 1974) is 578.69 million Sm^3 (or equivalent to 3.6 Giga barrel).

2.3.2 Exploratory effort, efficiency, and crude oil price

Turning to the relationship between exploratory effort, efficiency, and crude oil price. Two separate equations are estimated for waiting time (wt) and discovery size (s), with each equation containing the crude oil price (op) and the total of exploration wells (w). The hypothesis is that the waiting time between discoveries has a negative relationship with the oil price, while discovery size has a positive relationship with the oil price. The high oil price motivates more exploratory effort, proxied by more exploration wells, which is associated with a shorter time between discoveries. The high oil price also stimulates producers to find oil in a frontier area that, unlike a mature area, offers the prospect of making a larger size discovery. However, as discoveries become more frequent, the depletion rate becomes more associated with a smaller discovery size.

This subsection employs the yearly data from 1968 to 2019 for empirical analysis, and the discovery waiting time and size values are transformed into average annual data. The number of exploration wells is obtained from the Norwegian Petroleum Directorate (NPD) database.[2] The crude oil price data is available on the Energy Information Administration website.[3] The real oil price is applied by adjusting the nominal oil price with the U.S. Consumer Price Index (CPI) to account for global inflation. The CPI series are obtained from the U.S. Bureau of Labor Statistics and available from the U.S. EIA.[4] As for the robustness test, the purchasing power parities (PPP) are applied as an adjustment factor of nominal oil price to deal with the exchange rate fluctuations for non-US fluctuations. PPP is obtained from the OECD stat and measured in Norwegian Krone per U.S. dollar.[5]

The correlation between the waiting time and other exploratory effort variables are illustrated in Figures 2.6-2.7 and Table 2.2. For correlation and time-series plot purposes, this subsection applies yearly data from 1995 to 2020. The oil rig counts data for Norway are available on Baker Hughes website; these are also averaged yearly.[6] Exploration costs annual data are obtained from the Oil and Gas Department of Norway's Ministry of Petroleum and Energy.[7]

[2]https://factpages.npd.no/en/wellbore/Statistics/EntryYear
[3]https://www.eia.gov/outlooks/steo/realprices/
[4]https://www.bls.gov/cpi/data.htm
[5]https://data.oecd.org/conversion/purchasing-power-parities-ppp.htm
[6]https://rigcount.bakerhughes.com/intl-rig-count
[7]The exploration cost data was requested to the corresponding email to $OED -$

However, due to insufficient observations of rig counts and exploration costs time series, these are not included in the empirical regression. Table 2.2 shows that the NCS average waiting times have negative correlations with discovery size and other exploratory efforts, namely the number of exploration wells, rig counts, exploration costs, and crude oil price. Meanwhile, the crude oil price has a positive correlation with discovery size and the other exploratory efforts.

Table 2.2: Pairwise correlation between waiting time, other exploratory efforts, and crude oil price

	waiting time	size	well	rig	cost	oil price
waiting time	1					
size	-0.341	1				
well	-0.764	0.195	1			
rig	-0.147	-0.011	0.096	1		
cost	-0.755	0.277	0.811	0.076	1	
oil price	-0.581	0.372	0.530	0.179	0.691	1

Figures 2.6 and 2.7 support the evidence that in most periods, waiting time moves in an opposite direction to the oil price and the number of exploration wells. Figure 2.6 shows that when crude oil prices rise between 1995 and 2008, the waiting time between discovery decreases. After 2011, particularly between 2015 and 2016 and also after 2019, crude oil prices tend to drop. In these periods, the waiting time fluctuates. It then increases after 2019. Norwegian oil rig counts show an increasing trend from 1995 to 2001, then a decreasing trend up to 2008, as shown in Figure 2.7. In these periods, the average waiting times show an opposite trend; they decrease from 1995 to 2001 and increase from 2001 to 2008. After 2008, rig counts increase, and waiting times decrease. Then both remain relatively level with just slight fluctuations during 2009-2019. After 2019, the Norwegian oil rig counts decrease and waiting times increase. Figure 2.7 illustrates a more straightforward interpretation, in that the number of oil exploration wells has an increasing trend from 1995 to 2009, while the waiting times decrease. The waiting time fluctuates a little between 2010 and 2019, whereas exploration wells show a decreasing trend after 2015. In 2020,

Fakta@oed.dep.no and received on 12 February 2021 the number of exploration wells stayed low while the waiting time increases.

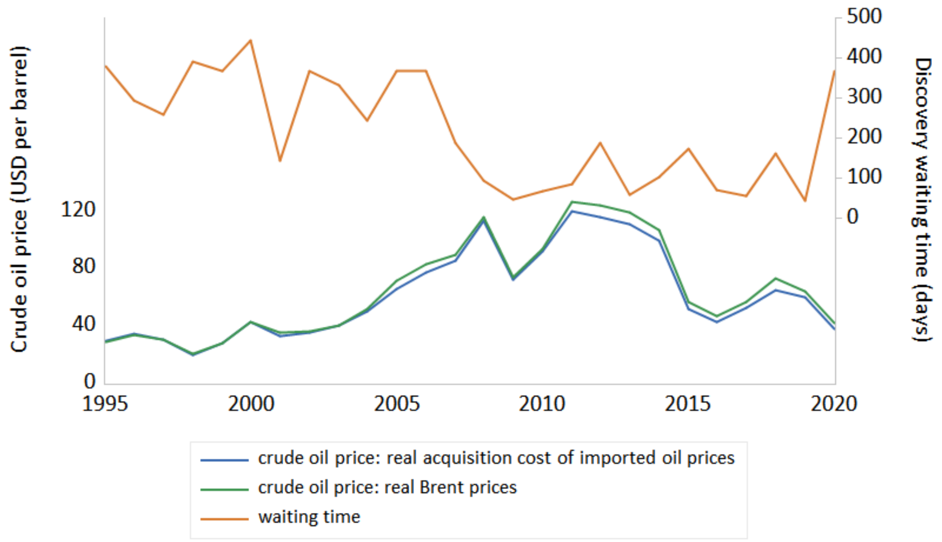


Figure 2.6: The Norwegian oil discovery waiting time and crude oil price

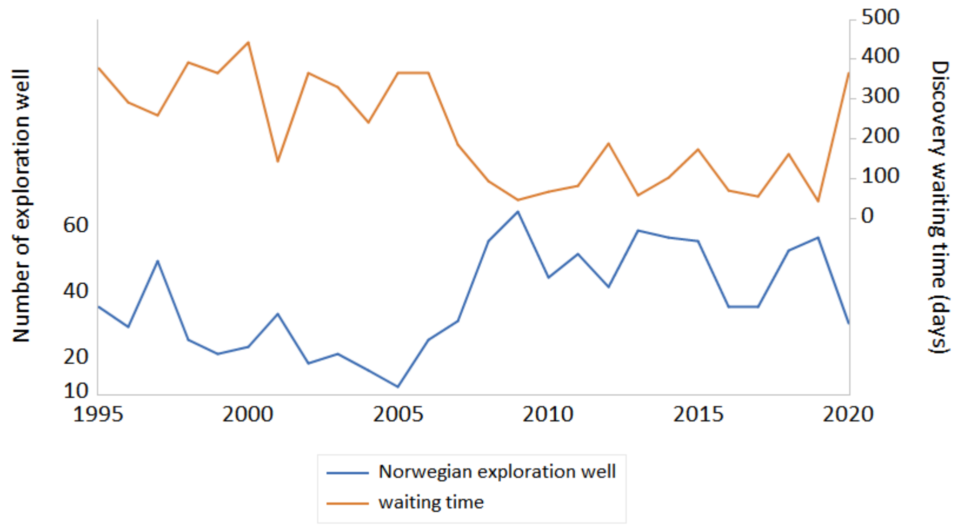


Figure 2.7: The Norwegian oil discovery waiting time and exploration well

2.4 Methodology

This chapter is organized into two main parts. The first parts deals with the simulation, setting out the tests run to fit the distribution type. The second part is empirical, in which the effect of oil price on exploratory effort and efficiency is estimated.

2.4.1 Monte Carlo simulation

Monte Carlo simulation aims to emulate a complex system that is difficult to practically and physically replicate by constructing a mathematical model (Thomopoulos, 2013). The essential principle of Monte Carlo simulation is applying a large number of random sampling. First, each input variable requires random variates. Second, a computation is run based on the pre-defined model. Each random input yields random outcomes. Third, the steps are repeated thousands of times to obtain thousands of outcomes. The flow requires a large number of random inputs to generate a large number of samples. This subsection utilises R software to run the Monte Carlo simulation in the following few steps:

1. Generate a random number between 0 and 1 for 141 sequences following the best fit distribution to obtain:
    - the discovery size expected values
    - the discovery time between one and subsequent discovery (waiting time) expected values
2. Repeat the process for 10,000 times of replication for each distribution.
3. To obtain the time distribution of each discovery, set the initial date as the first discovery date, then add the expected waiting time as generated by step 1.

4. Combine the expected value of discovery size with the expected waiting time value, according to their respective matrices location.

5. The simulation can also be applied when there is no historical data available by giving an estimate of the parameter. Specify the new parameter estimates for each distribution.

6. Repeat step 1 to step 4.

7. Test the goodness-of-fit statistic of the new simulated discovery size and waiting time by applying a Kolmogorov-Smirnov test. The test aims to test whether the simulated data have a distribution that is similar to the actual data.

Step 1 to 4 constitute the base scenario that employs the parameters fitted to the actual data. Steps 5 to 7 offer flexibility in allowing the simulation to be applied when there is no data available. However, these steps are also beneficial for sensitivity analysis to test the practicability of using waiting time as a proxy for exploratory efficiency to capture the effect of exploratory effort.

The probability distribution of the discovery size and waiting time is selected based on the smallest value of information criteria among a few candidates of probability distributions. After choosing the distribution, the parameter values need to be estimated. Thomopoulos (2013) emphasises four critical step of preliminary tests to be met by each random variable in the simulation: test for independence to avoid autocorrelation, calculate the statistical measures, choose the candidate of probability distributions, estimate parameters for the selected distribution, and test the goodness-of-fit.

2.4.1.1 Independence test

The aim of testing the independence of the sample is to obtain valid estimates. This subsection applies the BDS independence test proposed by

Broock et al. (1996). The test considers a pair of points and a distance ϵ that has a benefit of being invariant to the underlying series distribution. The probability of distance $c_1(\epsilon)$ being less than or equal to distance ϵ is identical for any points. For multiple pairs of points, the joint probability of every pair of points is $c_m(\epsilon)$ with m being the dimension, i.e., the number of consecutive data points applied. The assumption of independence is when $c_m(\epsilon) = c_1^m(\epsilon)$.

The probability is defined by the ratio of the number of sets that satisfies the ϵ condition and the total number of sets. For n observations of the data series x, the statistics of the correlation integrals $c_{m,n}$ is the following,

$$c_{m,n}(\epsilon) = \frac{2}{(n-m+1)(n-m)} \sum_{s=1}^{n-m+1} \sum_{t=s+1}^{n-m+1} \prod_{j=0}^{m-1} I_\epsilon\big(x_{s+j}, x_{t+j}\big) \tag{2.1}$$

where subscript s and t are observation points that create sets of pairs, and indicator function I_ϵ is

$$I_\epsilon(x,y) = \begin{cases} 1, & \text{if } |x-y| \leq \epsilon \\ 0, & \text{otherwise } 0 \end{cases}$$

Then, the test statistic is the following.

$$b_{m,n}(\epsilon) = c_{m,n}(\epsilon) - c_{1,n-m+1}(\epsilon)^m \tag{2.2}$$

The null hypothesis is that the observations of the series are independent and close to zero. The maximum dimension applied in this study is three, and bootstrap probability is used to give more accuracy in a small sample through a large number of repetitions.

2.4.1.2 Statistical measures

Statistical measures that are useful for selecting the probability distributions are minimum, maximum, mean, median, standard deviation, skewness, kurtosis, and quartile. The following equations define skewness and kurtosis (Casella and Berger, 2021):

$$sk(x) = \frac{E\left[\left(x - E(x)\right)^3\right]}{var(x)^{\frac{3}{2}}}, with\, unbiased\, estimator\, \hat{sk} = \frac{\sqrt{n(n-1)}}{n-2}\frac{m_3}{m_2^{\frac{3}{2}}} \tag{2.3}$$

$$kr(x) = \frac{E\left[\left(x - E(x)\right)^4\right]}{var(x)^2}, \hat{kr} = \frac{(n-1)}{(n-2)(n-3)}\left((n+1)\frac{m_4}{m_2^2} - 3(n-1)\right) + 3 \tag{2.4}$$

where m_k is given by $m_k = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^k$, and where m_2,m_3, and m_4 are empirical moments with n number of observations and $\bar{x}$ is the mean value of variable x.

2.4.1.3 The candidate of continuous probability distributions and parameter estimates

The parametric distributions defined in Figures 2.10 and 2.11 follow a distribution that is closer to four distributions of the observations: lognormal, Weibull, gamma, and exponential, as shown by skewness and kurtosis plots in Figures 2.8 and 2.9.

1. Lognormal

   Suppose x is a variable with a lognormal distribution and y is a variable with normal distribution, their relation would be the following:

$$y = ln(x); x = e^y \tag{2.5}$$

$$x \sim ln(\mu_y, \sigma_y^2) \tag{2.6}$$

$$y \sim N(\mu_y, \sigma_y^2) \tag{2.7}$$

$$\mu_y = ln\Big[\frac{\mu_x^2}{\sqrt{\mu_x^2 + \sigma_x^2}}\Big] \tag{2.8}$$

$$\sigma_y^2 = ln\Big[1 + \frac{\sigma_x^2}{\mu_x^2}\Big] \tag{2.9}$$

where ln denotes natural logarithm (Thomopoulos, 2013).

2. Gamma

The density for Gamma distribution is defined by the following.

$$f(x) = \frac{x^{k-1}\theta^k e^{-\theta x}}{\Gamma(k)};\ x \geq 0 \tag{2.10}$$

where $\Gamma(k)$ is the Gamma function and is

$$\Gamma(k) = \int_0^\infty t^{k-1}e^{-t}dt;\ for\ k > 0, \tag{2.11}$$

and the mean and variance are the following:

$$\mu = \frac{k}{\theta} \tag{2.12}$$

$$\sigma^2 = \frac{k}{\theta^2} \tag{2.13}$$

A random variate for Gamma distribution depends on whether the value is $k < 1$ or $k > 1$.

3. Weibull

The Weibull density is defined by the following equation with two parameters k_1 and k_2.

$$f(x) = k_1 k_2^{-k_1} x^{k_1} exp\Big[-\Big(\frac{x}{k_2}\Big)^{k_1}\Big];\ x > 0 \tag{2.14}$$

The cumulative distribution function is given by

$$F(x) = 1 - exp\Big[- \Big(\frac{x}{k_2}\Big)^{k_1}\Big];\ x > 0 \tag{2.15}$$

and the expected value and variance are the following.

$$E(x) = \frac{k_2}{k_1}\Gamma\Big(\frac{1}{k_1}\Big) \tag{2.16}$$

$$V(x) = \frac{k_2^2}{k_1}\Big[2\Gamma\Big(\frac{2}{k_1}\Big) - \frac{1}{k_1}\Gamma\Big(\frac{1}{k_1}\Big)^2\Big] \tag{2.17}$$

Obtaining a random uniform variate from $u \sim U(0,1)$ and setting it to $F(x)$, results in a random x from the Weibull distribution.

$$x = k_2\Big[- ln(1-u)\Big]^{\frac{1}{k_1}} \tag{2.18}$$

4. Exponential

As x decreases, the probability density $f(x)$ also decreases. The largest value is at $x = 0$.

$$f(x) = \theta e^{-\theta x},\ for\ x \geq 0 \tag{2.19}$$

The cumulative distribution function $F(x)$ is the following.

$$F(x) = 1 - e^{-\theta x},\ for\ x \geq 0 \tag{2.20}$$

The mean and variance are shown in Equation 2.21 and 2.22.

$$\mu = \frac{1}{\theta} \tag{2.21}$$

$$\sigma^2 = \frac{1}{\theta^2} \tag{2.22}$$

Then, obtaining a random variate from $u \sim U(0,1)$ and setting it to be equal to F(x), the random variate of x is the following:

$$x = -\frac{1}{\theta} ln(1-u) \tag{2.23}$$

Related to other distributions, the density shape is exponential when the parameter from Weibull distribution $k_1 \leq 1$ or when the parameter from Gamma distribution $k = 1$.

2.4.1.4 Model selection by information criteria

Information criteria are used to determine the model selection by measuring the distance of the model specification from the true model (goodness-of-fit). The commonly used information criteria are Akaike Info Criterion (AIC) and Schwarz Criterion (SC). The AIC and SC are based on the log-likelihood function and are adjusted by different penalty functions, in which SC is more parsimonious than AIC. Equations 2.26 and 2.27 describe the AIC and SC, respectively.

$$\left|\hat{\sum} \epsilon\right| = det\left(\frac{1}{T-(pk+d)}\sum_t \hat{\epsilon_t}\hat{\epsilon_t'}\right) \tag{2.24}$$

$$l = \frac{-T}{2}k(1+log(2\pi)) + log\left|\hat{\sum} \epsilon\right| \tag{2.25}$$

$$AIC = \frac{-2l}{T} + \frac{2N}{T} \tag{2.26}$$

$$SC = \frac{-2l}{T} + \frac{Nlog(T)}{T} \tag{2.27}$$

where T is the number of observation, p is the number of the lags, k is the number of the endogenous variables, d is the number of the exogenous variables (e.g., intercept), N is the total number of parameter estimated in all equations, $N = k(pk+d)$, $\hat{\Sigma}_\epsilon$ is the determinant of the residual covariance, and l is the log of the likelihood function. The optimal model chosen is the one that has the smallest value of information criteria.

2.4.1.5 The goodness-of-fit statistics

Goodness-of-fit is applied to the Norwegian oil discovery size and waiting time data to measure the distance between the fitted parametric distribution and the empirical distribution function. This subsection employs Kolmogorov-Smirnov (KS), Cramer-von Mises (CvM), and Anderson-Darling (AD) statistics defined by D'Agostino and Stephens (1986). [8]

The KS test assumes that the distribution parameters are known, while CvM and AD tests do not consider the *a priori* knowledge of the distribution parameters. CvM and AD tests estimate the parameters from the data by maximum likelihood. The null hypothesis H_0 and the alternative H_1 of each test are as follows:

H_0: The data comes from the specified distribution.

H_1: At least one value does not match the specified distribution.

The following formulas describe the computational formula for each test.

1. Kolmogorov Smirnov (KS) statistic

   $\max(D^+, D^-)$ with

$$D^+ = \max_{i=1,\dots,n}\Big(\frac{i}{n} - F(x_i)\Big), D^- = \max_{i=1,\dots,n}\Big(F(x_i) - \frac{i-1}{n}\Big) \tag{2.28}$$

2. Cramer-von Mises (CvM) statistic

$$CvM = \frac{1}{12n} + \sum_{i=1}^{n}\Big(F(x_i) - \frac{2i-1}{2n}\Big)^2 \tag{2.29}$$

[8]The tests are generated in R-software utilising the *fitdistrplus* R-package proposed by Delignette-Muller et al. (2015)

3. Anderson-Darling (AD) statistic

$$AD = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)[lnF\big(x_i\big) + ln\big(1 - F(x_{n-i+1})\big)] \qquad (2.30)$$

where F is theoretical cumulative distribution to be tested, which must be continuous distribution, $F(x)$ is the Cumulative Distribution Function (CDF) for the specified distribution, D is test statistic, n is sample size, and i is the -*th* sample with data sorted in ascending order.

2.4.1.6 Autoregressive Moving Average (ARMA) forecast model

The basic ARMA model is applied to compare the error between the forecast and simulation results. The first step is to consider an autoregressive equation with the p-th order, AR(p),

$$Y_t = \rho_0 + \rho_1 Y_{t-1} + ... + \rho_p Y_{t-p} + x_t$$

where x_t is a white-noise process following the MA(q) process and is expressed by the following equation:

$$x_t = \sum_{i=0}^{q}\theta_i \epsilon_{t-i}, \qquad (2.31)$$

The ARMA(p, q) equation is obtained by combining the moving average MA(q) with the autoregressive AR(p). The ARMA(p, q) equation is described in Equation 2.32, where p is the order of autoregressive and the q is the order of moving average (Enders, 2015).

$$Y_t = \rho_0 + \sum_{i=1}^{p}\rho_i Y_{t-i} + \sum_{i=0}^{q}\theta_i \epsilon_{t-i} \qquad (2.32)$$

2.4.1.7 Evaluation test

The predicted value of the numeric vectors for discovery size and waiting time are evaluated by measuring the differences from the actual values through the Mean Absolute Error (MAE) test. Equation 2.33 describes the formula of the MAE test.

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t| \tag{2.33}$$

2.4.2 Empirical framework

2.4.2.1 Unit root test

A unit root test is carried out to examine the stationarity of the series whether or not the variables are all integrated in the same order. A weak stationarity condition occurs when the mean and autocovariance do not depend on the time. The unit root test is important because standard inference cannot be applied when any of the regression series are integrated. The unit root test result is useful to determine the suitable empirical model, for instance, the Auto-Regressive Distributed Lag (ARDL) model is suitable when there are mixed of stationary and nonstationary variables. Vector autoregression (VAR) requires all variables to be stationary, while cointegrated VAR requires all to be integrated of the same order. This chapter applies the standard Dickey Fuller (DF) unit root test proposed by Dickey and Fuller (1979).

2.4.2.2 Structural break test

This chapter applies three structural break tests: supremum Wald, exponential Wald, and recursive. Andrews (1993) proposes the test for a structural break by testing the stability of the coefficients over time. This structural break applies the supremum Wald tests that utilises the maximum sample of the test.[9] Suppose π is the break-point and lies within the range of β_1 and β_2, the Supremum and exponential Wald test statistics W_T are given by

$$Supremum\ W_T = \sup_{\beta_1 \leq \pi \leq \beta_2} W_T(\pi) \tag{2.34}$$

$$Exponential\ W_T = ln\left[\frac{1}{\pi_2 - \pi_1 + 1}\sum_{\pi=\pi_1}^{\pi_2} exp\left(\frac{1}{2}W_T(\pi)\right)\right] \tag{2.35}$$

The null hypothesis is no break or the stability of the parameter,

$$H_0 : \beta_t = \beta_0 \tag{2.36}$$

and the alternative hypothesis is

$$H_1 : \beta_t = \begin{cases} \beta_1(\pi)\ for\ t = 1, .., T_\pi \\ \beta_2(\pi)\ for\ t = T_{\pi+1}, ..., T \end{cases} \tag{2.37}$$

The limiting distributions of the test statistics are

$$Supremum\ W_T \rightarrow_d \sup_{\lambda \in [\varepsilon_1, \varepsilon_2]} W(\lambda) \tag{2.38}$$

$$Exponential\ W_T \rightarrow_d\ ln\left[\frac{1}{\varepsilon_2 - \varepsilon_1}\int_{\varepsilon_1}^{\varepsilon_2} exp\left(\frac{1}{2}W(\lambda)\ d\lambda\right)\right] \tag{2.39}$$

$$W(\lambda) = \frac{\big(B_k(\lambda) - \lambda B_k(1)\big)'\big(B_k(\lambda) - \lambda B_k(1)\big)}{\lambda(1-\lambda)} \tag{2.40}$$

where, $B_k(\lambda)$ is a vector of k-dimensional independent Brownian motions, $\varepsilon_1 = \beta_1/T$, $\varepsilon_2 = \beta_2/T$, and $\lambda = \varepsilon_2(1-\varepsilon_1)/\{\varepsilon_1(1-\varepsilon_2)\}$.

[9] The test is documented in Stata using *estat sbsingle* command (StataCorp, 2017).

2.4.2.3 Autoregressive Distributed Lag (ARDL)

This chapter applies the ARDL approach to examine the relationship between time between one and subsequent discoveries and the oil market variables. The benefit of applying ARDL is its capacity to deal with a mixed order of integration among variables. Inference can be drawn for mixed integration and also without knowing whether the variables are integrated of order zero $I(0)$ or one $I(1)$. ARDL will not be applicable when any variable integrated of order 2, $I(2)$, as it causes spurious estimates. Hence, the unit root test is still important to carry out to ensure there is no variable integrated of order 2, $I(2)$, in the ARDL estimates.

ARDL time series consider the relationship of the contemporaneous and lagged values for both dependent and independent variables. The general ARDL model is expressed as Equation 2.41.

$$y_t = a_0 + a_1 t + \sum_{i=1}^{p} \phi_i \, y_{t-i} + \sum_{j=1}^{k} \sum_{i=0}^{q} \beta'_{j,i} \, x_{j,t-i} + \epsilon_t \tag{2.41}$$

The y_t is the dependent variable, $x_1, ..., x_k$ are k explanatory variables, with p and q being the lags of the dependent and explanatory variables, respectively. This study investigates the relationship between exploratory effort, efficiency, and crude oil price. Two equations are estimated with waiting time between discoveries (wt) and discovery size (s) as the dependent variable y_t in the first and second equation, respectively. The $x_1, ..., x_k$ is the number of exploration wells (w) and crude oil price (op) as the regressors. The linear trend t is assumed to be zero because the graphical interpretation of the series does not exhibit any trend.

The optimal lags p and q are obtained by minimising the value of information criteria AIC or SC, as described in Equations 2.26 and 2.27. The constant term is a_0, the error term is ϵ_t, and the respective coefficients a_1, ψ_i, β_{j,l_j} are the coefficients of the linear trend t, lags of y_t, and lags of the k explanatory variables $x_{j,t}$.

Table 2.3: ARDL model deterministic cases

Case Number	Description
1	no constant, no trend
2	restricted constant, no trend
3	unrestricted constant, no trend
4	unrestricted constant, restricted trend
5	unrestricted constant, unrestricted trend

2.4.2.4 Autoregressive Distributed Lags in Error-Correction form (ARDL-EC) and ARDL with a break

The basic ARDL equation follows Equation 2.41 with two explanatory variables ($k = 2$): exploration well and real oil price. The empirical study in this chapter runs two single ARDL equations for waiting time and discovery size as dependent variables. In the first ARDL equation, y_t is the waiting time between discoveries as the dependent variable, and x1, x2 are the number of exploration wells and crude oil prices as the explanatory variables. The second ARDL equation applies discovery size as y_t, with x_1, x_2 being exploration wells and the oil price.

ARDL-EC is a way of parameterising the regressors in the long-run relationship. The bounds testing by Pesaran et al. (2001) is a post-estimation regression to examine the existence of the long-run equilibrium relationship without specifying whether the order of integration for each variable is integrated order zero $I(0)$ or one $I(1)$. ARDL-EC allows the relationship between the variables in the long run to be pure $I(0)$, $I(1)$, or cointegrated, which refers to the condition when the linear combination of nonstationary variables becomes stationary in the long run. The two equations follow case three from Table 2.3, which is unrestricted constant and no trend. The dependent variable is in the first difference form and Equation 2.41 is split into three main components:

speed-of-adjustment, long-run coefficients, and short-run coefficients. Equation 2.42 describes the ARDL-EC parameterisation based on Engle and Granger (1987); Hassler and Wolters (2006); Kripfganz et al. (2018).

$$\Delta y_t = a_0 + a_1 t - \alpha(y_{t-1} - \theta x_{t-1}) + \sum_{i=1}^{p-1} \psi_{yi}\, \Delta y_{t-i} + \omega^{'} \Delta x_t + \sum_{i=1}^{q-1} \beta^{'}_{xi}\, \Delta x_{t-i} + \epsilon_t \tag{2.42}$$

The speed-of-adjustment α represents how much the dependent variable needs to adjust from the previous period to revert to equilibrium in the current period, and it is given by $\alpha = 1 - \sum_{j=1}^{p} \phi_i$ where ϕ refers to the coefficient of the lagged dependent variable in Equation 2.41. As the test distribution is not standard, the critical values of the speed-of-adjustment follow Kripfganz et al. (2018). The long-run coefficients θ is given by $\theta = \frac{\sum_{j=0}^{q} \beta_j}{\alpha}$ and represents the effects of the independent variables on the dependent variable in equilibrium. [10] The short-run coefficients, ψ, $\omega^{'}$, $\beta^{'}$ indicate short-term fluctuations of the explanatory variables on the dependent variables.

The long-run equilibrium or cointegration evidence requires both F-statistics and t-statistics to be rejected. The bounds testing calculates the F-statistics of the joint null hypothesis that the speed of adjustment and the coefficients of the lagged explanatory variables are zero $H_0^F : (\alpha = 0)\big(\sum_{j=0}^{q} \beta_j = 0\big)$. The F-statistics are then compared to the critical values. This chapter follows Kripfganz et al. (2018) for the finite-sample and asymptotic critical values that provide more efficient estimates, regardless of the lag length or the number of short-run coefficients. If the F-statistics is rejected, then the value of t-statistics needs to be checked. The null hypothesis is that the coefficient of the speed of adjustment is zero $H_0^t : \alpha = 0$.

Chapter 2 considers the ARDL model with a structural break as the data

[10] The Stata command *estat ectest* proposed by Kripfganz et al. (2018) is applied to obtain the coefficient estimates of ARDL-EC

series contain a few shocks mainly due to oil price fluctuation. As the time series applies annual data, which results in a relatively small sample size of 53, the model takes into account a break in intercept only to avoid overfitting issues when there are too many regressors in a small sample size. The model with a structural break takes the following form in Equations 2.43 and 2.44 for the basic ARDL and ARDL-EC with the break, respectively.

$$y_t = a_0 + a_1 t + \sum_{i=1}^{p} \phi_i \, y_{t-i} + \sum_{j=1}^{k} \sum_{i=0}^{q} \beta^{'}_{j,i} \, x_{j,t-i} + \gamma_0 b_t + \epsilon_t \tag{2.43}$$

$$\begin{aligned} \Delta y_t = a_0 + a_1 t - \alpha(y_{t-1} - \theta x_{t-1}) + \sum_{i=1}^{p-1} \psi_{yi} \, \Delta y_{t-i} + \omega^{'} \Delta x_t \\ + \sum_{i=1}^{q-1} \beta^{'}_{xi} \, \Delta x_{t-i} + \gamma_0 b_t + \epsilon_t \end{aligned} \tag{2.44}$$

A dummy variable (b) represents the break in crude oil price, oil exploration well counts, and waiting time between discoveries in the first equation and oil price, well counts, and discovery size in the second equation. It takes a value of 1 from the estimated break date onwards and a value of 0 from the beginning of observation until before the estimated break date. The coefficient estimate of a break in intercept is denoted as γ_0. The estimated breakpoint for each equation is obtained based on the supremum Wald test by Andrews (1993).

2.5 Distribution type of NCS oil discovery size and time between discoveries

This section describes the properties of NCS discovery size and time between discoveries from which the underlying distribution data type can be assessed and a candidate selected for the distribution for the simulation. Figures 2.8 and 2.9 illustrate the plot of the square of skewness and the kurtosis of the observations as proposed by Cullen and Frey (1999), while empirical distribution plots help to fit the distribution set by the data . The plots are generated by utilising the package fitdistrplus in R (see Delignette-Muller et al. (2015) for more detail).

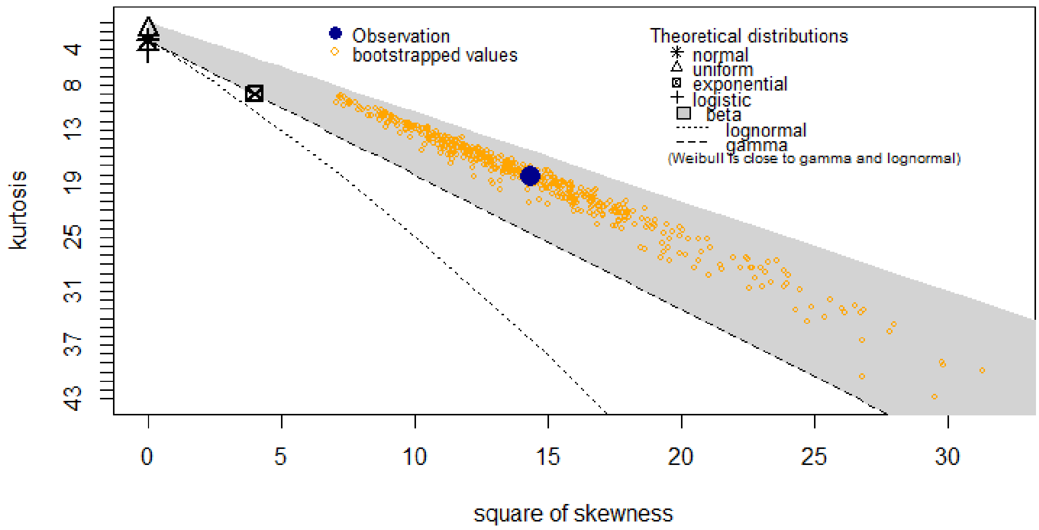


Figure 2.8: Skewness and Kurtosis of NCS oil discovery size

Figures 2.8 and 2.9 illustrate more clearly the skewness and kurtosis values mentioned in Table 2.1; these are useful to explain the distribution type. Skewness measures the symmetry of the data, while kurtosis identifies the fatness of the data set's tail. The skewness and kurtosis are respectively linked to third and fourth moments from the central value, and show how far the data is distributed relative to the benchmark of normal distribution. A

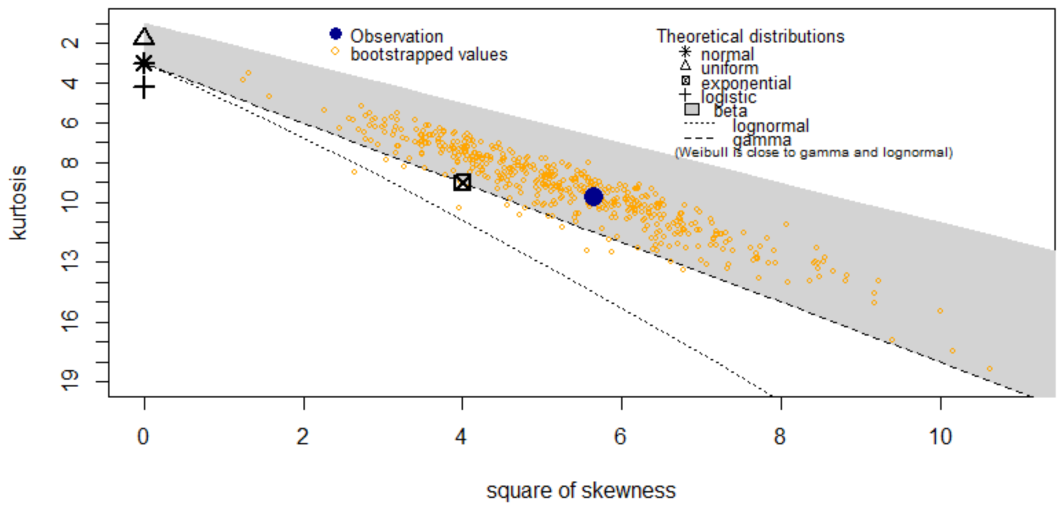


Figure 2.9: Skewness and Kurtosis of NCS time between discoveries

convention used in this scope is that a normal distribution has zero skewness and a kurtosis of three. Figures 2.8 and 2.9 also describe some predefined distributions (other than the normal distribution), namely uniform, logistic, exponential, gamma, and Weibull distributions. This is to narrow down the candidates for distribution selection. The possible distribution values for the data sets are shaded in a dark blue circle shaped by the yellow bootstrapped values.

A zero value of skewness implies symmetric empirical distribution, and it implies a non-symmetric distribution for non-zero skewness value. Both discovery size and waiting time data have non-zero positive values of skewness. That indicates that the data are asymmetric, skewed right, and have a right tail that is longer than the left. The high positive kurtosis (relative to the value of three) for both waiting time and size indicates the heavy-tailed distributions relative to the normal distribution.

Discovery size observations show that the distribution lying within lognormal, gamma, Weibull, and the time in between one and subsequent discovery observations are quite close to the exponential distribution. Hence, the predefined distributions adopted for the empirical and theoretical distribution plots

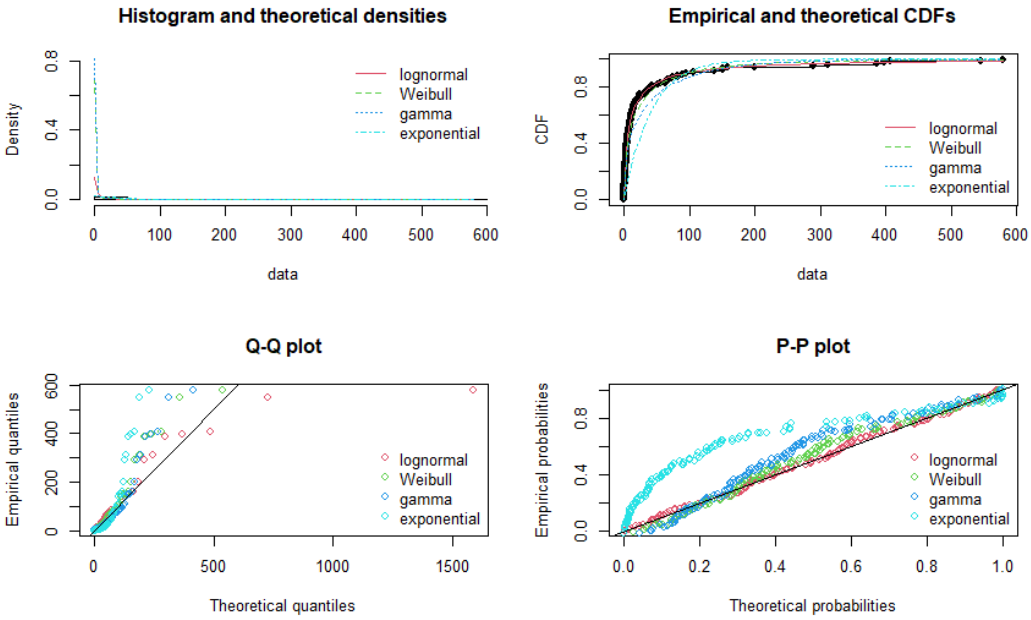


Figure 2.10: Empirical and theoretical NCS oil discovery size distribution

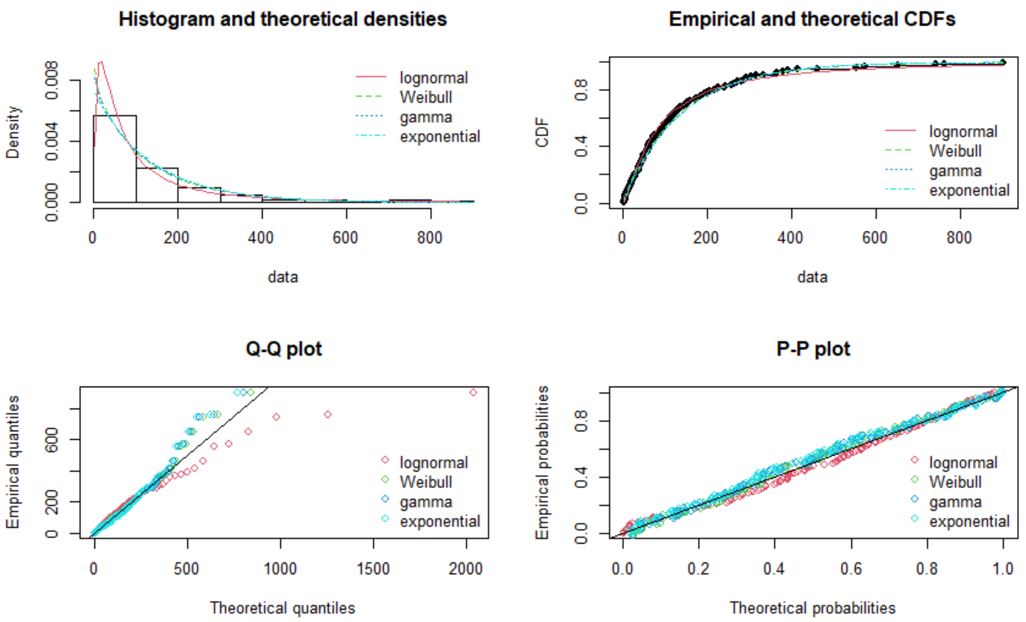


Figure 2.11: Empirical and theoretical NCS time between discoveries distribution

are Weibull, lognormal, gamma, and exponential, as shown in Figures 2.10 and 2.11. Figure 2.10 clearly illustrates a cumulative distributive function and empirical probabilities, indicating that lognormal and Weibull are close to the discovery size observation points. The exponential and Weibull distributions are close to the time in between discoveries observations, as shown by Figure 2.11. However, a further goodness-of-fit test must be employed to obtain the distribution with the best fit for discovery size and waiting time.

Table 2.4: Goodness-of-fit criteria for discovery size and time between discoveries distribution

Information Criteria	Lognormal	Weibull	Gamma	Exponential
NCS discovery size				
AIC	1158.183	1174.196	1198.729	1331.781
SIC	1164.081	1180.094	1204.627	1334.730
NCS discovery waiting time				
AIC	1653.161	1649.132	1649.838	1648.192
SIC	1659.030	1655.001	1655.707	1651.127

Table 2.4 summarises the information criteria of the predefined distributions for NCS discovery size and time between discoveries. Based on AIC and SIC, the smallest value for the information criteria is a lognormal distribution for discovery size, and the exponential distribution for time between one and subsequent discovery. Therefore, fur further analysis, this study considers lognormal and exponential distributions for size and time between discoveries, respectively, as the preferred distribution.

2.6 Simulation results and analysis

This section presents and analyses the Monte Carlo oil discovery simulation results (see Subsection 2.6.1) and the ARDL empirical analysis (Subsection 2.6.2). The NCS discovery sequence is adopted as the sample of observations illustrating discovery behaviour in the mature petroleum province. The simulation has certain objectives. First, it models the underlying distribution profile and forecasts the oil discovery sequences by simplifying the real system. Second, the simulation obtains and quantitatively understands the exploratory efficiency in the NCS mature petroleum province by identifying discovery size and time between one and subsequent discovery (i.e., waiting time) parameters under the same parent distribution. Third, the simulation captures the changes in the exploratory effort via parameter changes to understand the impacts of exploratory effort on waiting time and discovery size. The empirical analysis examines quantitatively how crude oil price fluctuation affects the waiting time and discovery size, drawing on the number and size of exploration wells to capture exploratory effort.

2.6.1 Simulation of discovery waiting time and size in a mature petroleum province

Subsection 2.6.1 focuses on the simulation results of the discovery sequences for the NCS mature petroleum province. This study simulates discovery size and the time between discoveries to simplify the real and complex oil discovery problem. As explained in Section 2.5, for best fit, the simulation follows a lognormal distribution for NCS discovery size and an exponential distribution for waiting time. There are three sets of discovery sequence simulations that aim to capture the changes in exploratory efforts. The first set of Monte Carlo simulations, which is the base scenario, is based on discovery size and the waiting time parameters generated by the 141 NCS discovery sequential sampling. The other two sets of simulations apply different parameters of discovery size and waiting time but with the same distribution as the NCS series. The second set of simulations applies parameter values that are lower than the base scenario's, while the third set uses parameter values that are higher than those of the base.

Before proceeding with the simulation for parameter changes, the independence of the series in the base scenario needs to be tested. Table 2.5 shows the BDS independence test for NCS actual discovery size and waiting time between discoveries. The BDS statistic cannot reject the null hypothesis that the series are independent. The test concludes that the size and waiting time data are independent. Five sets of simulated size and waiting time samplings are also tested for their BDS independence, and these results also confirm the independence of the simulated discovery sequences. Hence, the simulated data can be further analysed.

Table 2.5: BDS Independence test

Series	BDS Statistic	Bootstrap Probability
waiting time (wt)		
NCS data	0.015	0.215
Simulated data #1	0.037	0.079
Simulated data #2	0.021	0.314
Simulated data #3	0.029	0.186
Simulated data #4	-8e^{-4}	0.864
Simulated data #5	0.034	0.150
size (s)		
NCS data	-0.002	0.821
Simulated data #1	-0.022	0.686
Simulated data #2	-0.017	0.716
Simulated data #3	-0.024	0.341
Simulated data #4	0.018	0.456
Simulated data #5	-0.006	0.972

The base scenario reports the discovery size mean as 1.98 (in logarithmic value) with a standard deviation of 2.00; the exponential rate of waiting time between discoveries is 0.0073. Then, parameter changes are employed to capture changes in exploratory effort. For consistency of analysis, the high and low parameters of discovery mean and rate are determined by values that give the same distribution as the NCS discovery sequences (by trial error). The KS, AD, and CvM tests are applied to examine whether the simulated and actual NCS discovery size and waiting time sequences are from the same distribution. Those goodness-of-fit tests are applied to randomly taken samples (e.g., 1^{st}, 100^{th}, 500^{th}, $1,000^{th}$, $5,000^{th}$, and $10,000^{th}$ replication) for each set of the simulated sequence. The null hypothesis (that simulated and actual series are drawn from the same population) fails to be rejected at 5% significance level for most of the samples (for details, see Appendix A.1). Hence, the analysis to capture exploratory effort through parameter changes can proceed further.

Deciles help understand the probability of the true value of the random variables. As the sample is continuous data, the probability is given to the range of the discovery size and waiting time values. Tables 2.6-2.8 describe the three parameters and their decile values in three sets of discovery size and waiting time simulations. The simulated waiting time is the simulated lag time (in days) between one discovery and the discovery that follows it, which is the result of the exploratory effort. Summing simulated waiting time into an initial discovery date results in the simulated discovery date, which is useful for identifying the last date of discovery. Knowing the last discovery date makes it easier to understand the big picture of how exploratory effort affects discovery size and waiting time.

Table 2.6: Decile of simulated discovery size

log-mean	log-sd	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
1.7675	2.00	0.45	1.09	2.05	3.53	5.85	9.73	16.74	31.56	76.11
1.9801	2.00	0.56	1.34	2.53	4.35	7.22	11.96	20.59	38.83	93.67
2.3575	2.00	0.81	1.95	3.69	6.36	10.55	17.52	30.05	56.71	137.01

Table 2.7: Decile of simulated discovery waiting time

Rate	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
0.0068	15.41	32.71	52.31	74.96	101.75	134.59	176.81	236.65	338.73
0.0073	14.44	30.68	49.09	70.22	95.18	125.91	165.48	221.37	316.35
0.0090	11.66	24.79	39.66	56.73	77.04	101.82	133.71	178.54	255.32

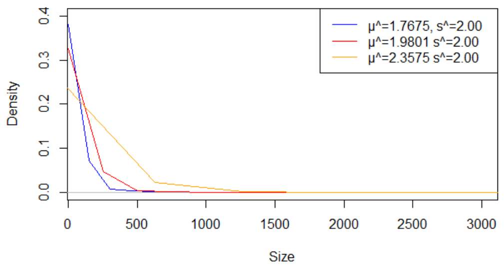
Table 2.8: Decile of simulated discovery dates

Rate	0.10	0.20	0.30	0.40	0.50
0.0068	30-01-1968	21-02-1980	16-09-1987	13-06-1996	16-10-2006
0.0073	17-01-1968	12-05-1979	19-06-1986	14-08-1994	04-04-2004
0.0090	06-01-1972	30-01-1977	28-10-1982	31-05-1989	02-04-1997

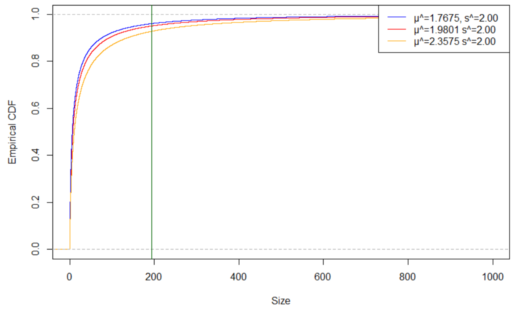
Rate	0.60	0.70	0.80	0.90
0.0068	20-06-2019	05-10-2035	10-11-2058	03-04-2098
0.0073	12-02-2016	23-05-2031	18-12-2052	12-08-2089
0.0090	26-10-2006	16-02-2019	06-06-2036	24-01-2066

The average NCS discovery size is in the upper quartile of the discovery size distribution, and the average waiting time is in the second quartile (the details are in Table 2.1). The simulated discovery size is also in the upper quartile. Applying the NCS parameter, the random variables of simulated discovery size have a 90% probability of being less than 93.67 million Sm^3. Meanwhile, 90% of the simulation result for the waiting time is less than 316 days, with a probability of the final discovery being made before August 2089. Taking the value of the giant oil field as greater than 500 million barrels [11]The standard size for giant discovery draws on Nehring (1978) and Zou (2013), and 500 million barrels is equivalent to 79.49 million Sm3, there is 90% probability that discovery size will be less than this.

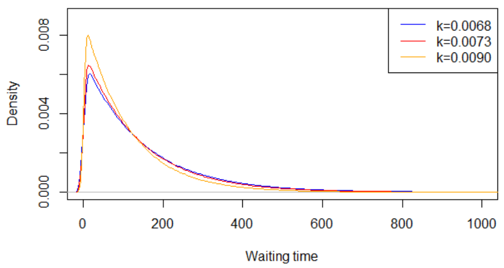
An interesting finding comes from the other waiting time and discovery date parameters. The 90% probability of simulated waiting time is less than 339 days for the low parameter, and less than 255 days for the high parameter. A low exponential rate leads to a longer waiting time between discoveries, and it causes a 90% probability of the final discovery being made before April 2098. In contrast, a high exponential rate results in a shorter waiting time and simulates 90% probability of the final discovery being before January 2066. The waiting time is an effect of the frequency of exploratory effort. To answer the question about the impact of the exploratory effort on waiting time, the more frequent or greater the exploratory effort, the shorter the time between one and subsequent discovery, and the shorter the waiting time (i.e., it has a lower value) that leads to final oil discovery.



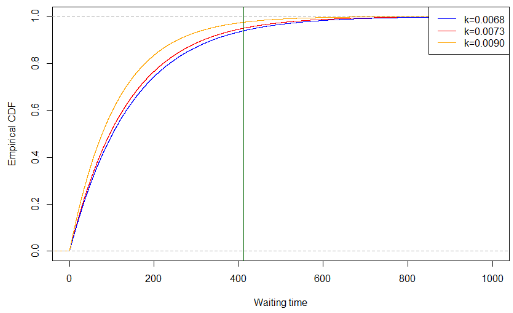
(a) Density of simulated discovery size



(b) Empirical Cumulative Distribution Function of simulated discovery size*



(c) Density of simulated waiting time



(d) Empirical Cumulative Distribution Function of simulated waiting time*

Figure 2.12: Density and Empirical Cumulative Distribution Function of simulated size and waiting time

*The vertical green line shows the 5% upper quantile.

Figures 2.12a and 2.12c show the density of simulated discovery size and waiting time. In Figure 2.12a, the standard deviations are kept constant for the three sets of simulations as the focus is on the discovery size changes. Meanwhile, the standard deviation changes represent the spread and affect the shape rather than the location of the density. As illustrated in Figure 2.12a, the larger the discovery size, the lower the probability density. The density figure shows that the probability of finding a smaller oil field size is higher than that of finding a bigger field. Figure 2.12b depicts the empirical cumulative distribution function for the simulated discovery size. The green line shows the 5% upper quantile or 95% confidence level of the NCS parameter-based simulation, i.e., 95% of the confidence level that the mean of discovery size lies below 194.13 million Sm^3.

Figure 2.12c shows that the higher the exponential rate, the higher the probability density. The probability density of a shorter waiting time is higher than that of a longer waiting time. The density figure shows more frequency (associated with more exploratory effort) for the higher exponential rate and the shorter waiting time. To sum up, the probability of making findings with a shorter time between one and subsequent discoveries and a smaller discovery size is high in the application of simulated NCS discovery. The 95% confidence level for NCS parameter-based simulation describes the true waiting time parameter between discoveries as lying below 411 days. Subsection 2.6.2 presents the estimation of empirical analysis, applying an ARDL framework to understand the quantitative relationship between exploratory effort, efficiency, and crude oil price.

Table 2.9: Mean Absolute Error (MAE) comparison of the Monte Carlo simulation and the ARMA(1,1) estimation

(a) NCS-based simulation

Variable	Actual mean	Simulated mean	Lower limit (5^{th} percentile)	Median (50^{th} percentile)	Upper limit (95^{th} percentile)
size	41.0806	53.1841	0.0007	0.0048	0.0354
waiting time	136.2357	137.3913	0.0072	0.0510	0.0365

(b) In-sample ARMA(1,1) forecast

Variable	Actual mean	Forecasted mean	Lower limit (5^{th} percentile)	Median (50^{th} percentile)	Upper limit (95^{th} percentile)
size	41.0806	45.7725	0.0654	0.1354	0.5213
waiting time	136.2357	156.5412	0.4091	0.2402	0.5896

This study applies the mean absolute error (MAE) test to evaluate the relative error between the Monte Carlo simulated data and the actual data. MAE calculates the deviation by the average difference between the simulated and actual discovery sizes and waiting times as an absolute value. The closer the value to zero, the more desirable the prediction (Abraham and Ledolter, 1983). However, there is no cut-off value from which one can conclude that the forecasted values are good or bad for a model. However, relative error is helpful in comparing the models. In this study, the Monte Carlo simulation error is compared with the in-sample ARMA forecast to assess the accuracy of the proposed simulation model.

Table 2.9 shows the MAE values taken from the NCS-based simulation with 10,000 replication and the in-sample ARMA(1,1) estimation. The MAE values for lower limit (5^{th} percentile), median (50^{th} percentile), and the upper limit (95 $^{\text{th}}$ percentile) are useful to evaluate the range of the average error of the simulation. The Monte Carlo simulation obtains lower values of MAE than those forecast by in-sample ARMA for the 5^{th}, 50^{th}, 95^{th} percentiles of discovery size and waiting time. In this case, the measure of accuracy is the simulated data based on NCS-parameters, thus: discovery size log-mean of 1.9801, log-standard deviation of 2, and exponential rate of 0.0073. The mean of the lower, median, and upper limit of the simulated data as compared to actual size and waiting time is described in Table 2.1. MAE values for discovery size simulation range from 0.07-3.5% and are from 0.72-5.10% for the waiting time between discoveries. The in-sample ARMA(1,1) forecast results in quite large error values of between 6.5-52% for discovery size and 24-59% error for waiting time estimation. This error comparison improves the confidence level of the true mean in the proposed Monte Carlo simulation for discovery size and waiting time between discoveries.

Combining the simulated discovery size and waiting time for each replication into a discovery sequence plot allows for a more straightforward interpretation of the exploratory effort in a mature petroleum province. Figures 2.13 - 2.15 illustrate the simulation of discovery sequence randomly taken from 1^{th}, 100^{th}, 500^{th}, $1{,}000^{\text{th}}$, $5{,}000^{\text{th}}$, and $10{,}000^{\text{th}}$ replications of NCS-based, low, and high parameters. The sequence of NCS-based parameters has few outliers for sizes above 200 million Sm^3. The low mean of average size parameter (shown in Figure 2.14) has fewer sizes above 200 million Sm^3 and the low exponential rate has final simulated discovery years that are further away than those of the

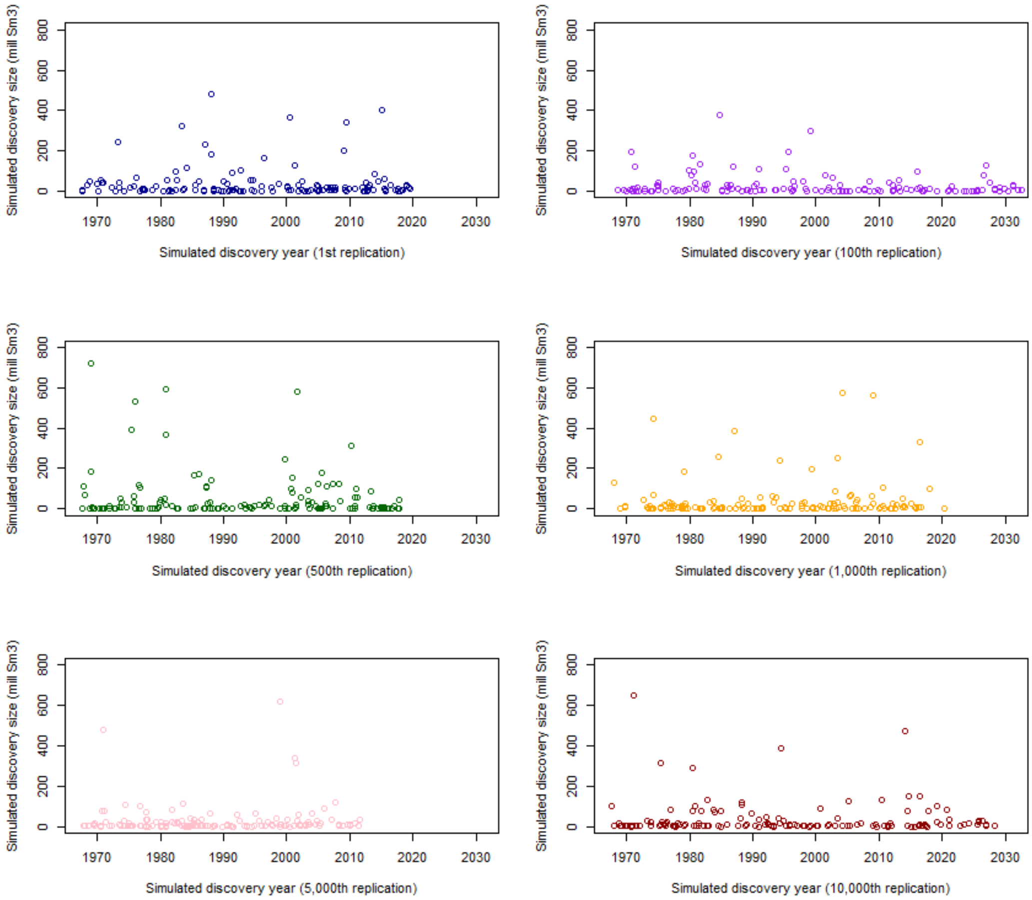


Figure 2.13: Simulated NCS discovery size and year

NCS parameters. In contrast, the simulation plots of the high mean of average size parameter, as illustrated in Figure 2.15, show more discovery sizes above 200 million Sm^3 than the NCS-based one, and the high exponential rate has a shorter last oil discovery year. To conclude, for high exponential rate, more exploratory effort leads to a reduction in waiting time between discoveries and causes a shorter time to the last oil discovery. The simulation exercises verify that the waiting time and discovery size are the results of exploratory effort.

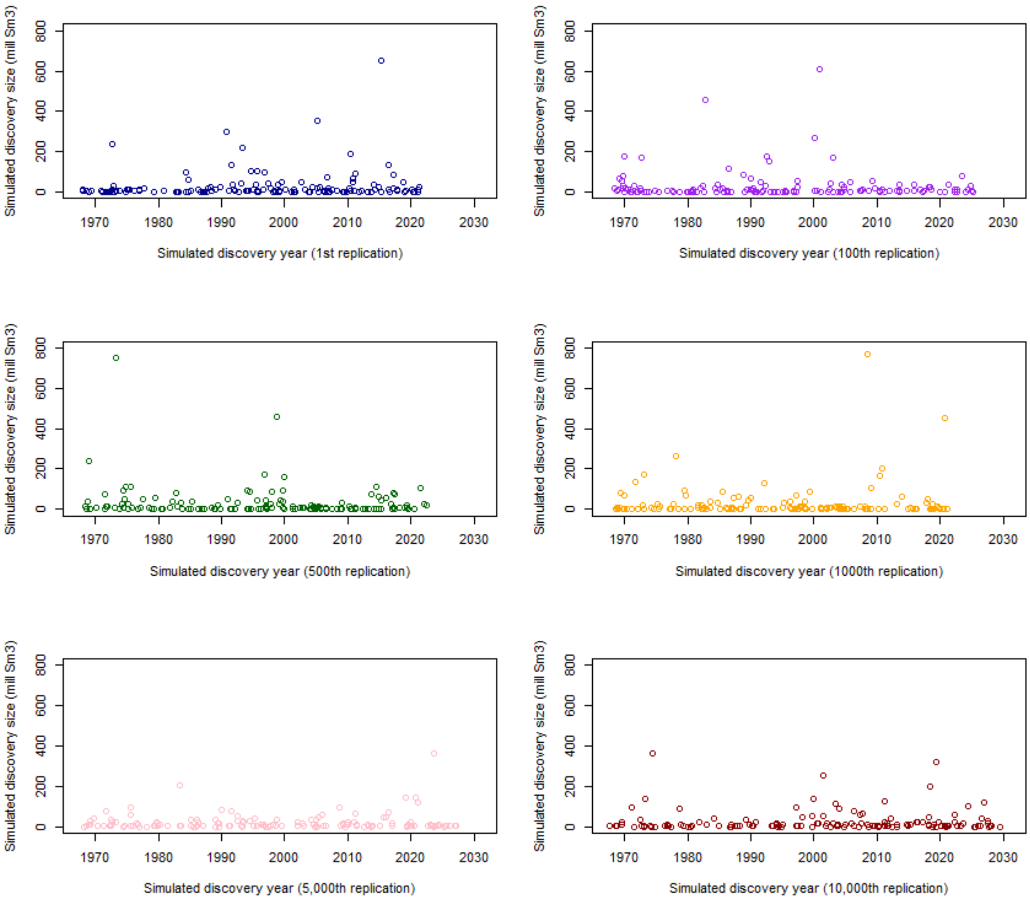


Figure 2.14: Simulated discovery size and year (low mean of discovery size and rate)

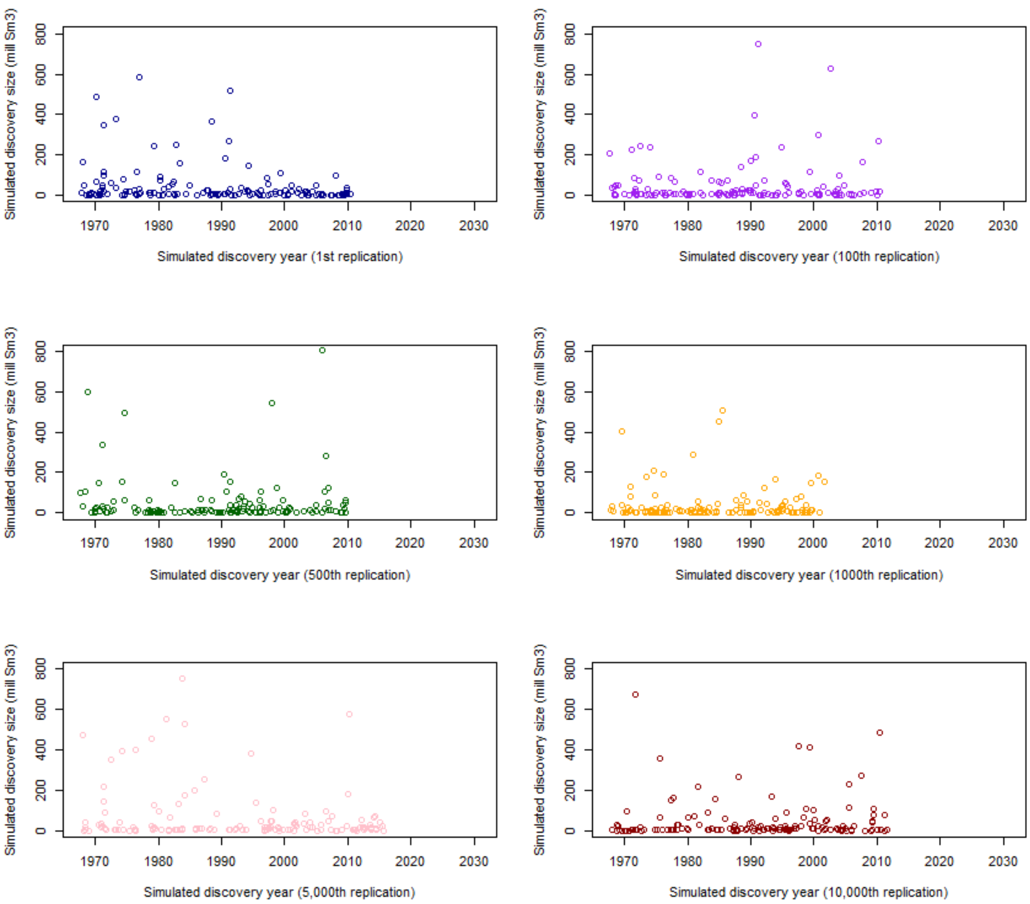


Figure 2.15: Simulated discovery size and year (high mean of discovery size and rate)

2.6.2 Empirical results of exploratory effort, efficiency, and crude oil price

The second part of the analysis discusses the relationship between exploratory effort and efficiency in NCS, as a mature petroleum province and the crude oil price. A careful examination investigates the effect of the crude oil price (op) changes on the waiting time between discoveries (wt) and discovery size (s), incorporating the number of oil exploration wells (w) as the proxy of exploratory effort; these are commonly used in the established literature.

One of this chapter's main contributions is to take into account the time between one and subsequent discovery as an alternative approach for reflecting exploratory efficiency. The difference between waiting time and existing measures such as exploration wells and rig counts is that the waiting time between discoveries is the result of exploratory effort, thus, a shorter waiting time is expected. In contrast, the exploration well count and rig count both directly proxy for exploratory effort. The more exploratory effort, the higher the number of exploration well and rig count, which leads to a shorter time between discoveries. The economic interpretation of the oil market is that a high crude oil price reflects more demand for crude oil at a given rate of supply. Increased demand boosts the number of exploration wells or rigs, causing a shorter waiting time between discoveries. Therefore, it is expected that the empirical estimation finds that waiting time has an opposite sign to that of the exploration wells and crude oil price.

Figures 2.16 and 2.17 respectively show the simulated waiting time taken from the NCS-based parameter simulation for two of the oil market variables, namely crude oil price and Norwegian oil exploration wells. The time series plot illustrates that simulated waiting time increases when real oil prices and the exploration well count drop after 2010. The waiting time results from the lag time in days between one discovery and another. For the graph and estimation

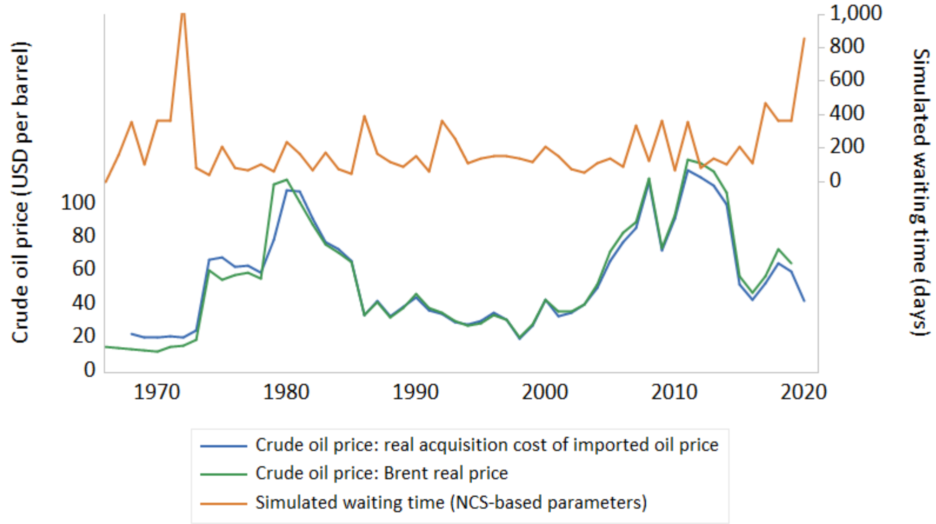


Figure 2.16: The simulated Norwegian oil discovery waiting time and crude oil price

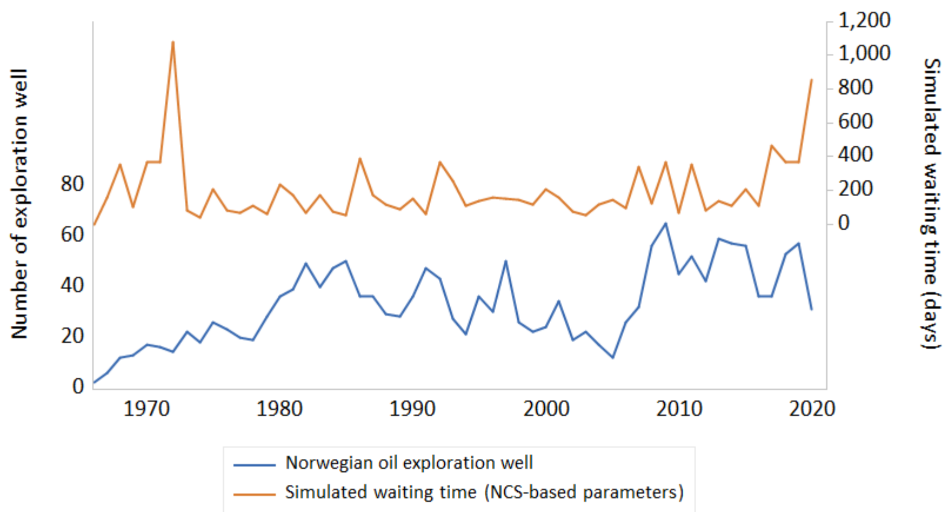


Figure 2.17: The simulated Norwegian oil discovery waiting time and oil exploration well

purposes, the waiting time series is averaged yearly. The high value of waiting time reflects the longer waiting time between discoveries. When the oil price is low, the number of exploration wells is also low, and it takes a long time between findings as there is no incentive to explore new fields. However, high oil price boosts new exploration, increasing the number of oil exploration wells and shortening the time between oil discovery.

Table 2.10: Unit Root Test for NCS discovery, exploratory effort, and crude oil price variables

Variable	ADF			
	Level	log-level	First difference	Log first difference
wt	-5.108***	-4.158***	-5.187***	-7.482***
s	-0.982	-3.880***	-3.656***	-3.842***
w	-3.226**	-2.187	-8.652***	-7.271***
op	-1.844	-1.345	-12.251***	-11.417***

Notes:
* $p < .10$, ** $p < .05$, *** $p < .01$
Variables expressed as wt is the waiting time between discoveries, s is the oil discovery size, w is the oil exploration well counts, and op is the real oil price.

Table 2.11: ARDL estimation for exploratory effort, efficiency, and crude oil price

	(i) wt_t	(ii) ls_t
Coefficient estimates ($\phi_i, \beta'_{j,i}$)		
wt_{t-1}	-0.007 (0.121)	
wt_{t-2}	0.221* (0.118)	
ls_{t-1}		0.580*** (0.164)
ls_{t-2}		-0.314* (0.180)
lw_t	-35.394 (57.375)	-2.355*** (0.763)
lw_{t-1}	-144.437** (62.790)	
lw_{t-2}	-27.050 (65.505)	
lw_{t-3}	139.993*** (47.527)	
lop_t	-53.947 (33.958)	0.568 (0.444)
(a_0) Constant	597.081** (242.909)	8.213*** (2.605)
Observation	50	50
F-statistic	5.68	6.70
Prob (F-statistic)	0.0001	0.0007

Notes:
Standard Error in parentheses; * $p < .10$, ** $p < .05$, *** $p < .01$
The notation "l" indicates natural logarithmic of the variable,
e.g. lw_t indicates logarithmic value of exploration wells at period t.

The unit root test is conducted to test the stationarity of the series and the results are shown in Table 2.10. The ADF test shows that there is mixed order

Table 2.12: ARDL estimation in error correction form for exploratory effort, efficiency, and crude oil price

	(i) $D.wt_t$	(ii) $D.ls_t$
(i) *Case 3*		
Bound Test H_0: no level relationship		
F-stat‡	7.404***	6.359**
t-stat‡	-4.280**	-4.073**
(ii) *Adjustment factor* *(-α)‡*		
wt_{t-1}	-0.786*** (0.184)	
ls_{t-1}		-0.734*** (0.180)
(iii) *Long-run* (θ)		
lw_{t-1}	-85.106 (56.610)	-3.209*** (0.957)
lop_{t-1}	-68.640* (37.432)	0.774 (0.561)
(iv) *short-run* ($\psi_{yi}, \omega^{'}, \psi^{'}_{xi}$)		
$D.wt_{t-1}$	-0.221* (0.118)	
$D.ls_{t-1}$		0.314* (0.180)
$D.lw_t$	-35.394 (57.375)	-2.355*** (0.763)
$D.lw_{t-1}$	-112.943* (60.474)	
$D.lw_{t-2}$	-139.993*** (47.527)	
$D.lop_t$	-53.947 (33.958)	0.568 (0.444)
(a_0) Constant	597.081** (242.909)	8.213*** (2.605)
Observations	50	50

Notes:
Standard error in parentheses; * $p < .10$, ** $p < .05$, *** $p < .01$
‡The approximate p-values applied for speed-of-adjustment coefficient is based on Kripfganz et al. (2018)

of stationarity of variables in level and log-level. The waiting time between discoveries is stationary at 5% significance level. In contrast, real oil price is not stationary at either level or log-level. Size is stationary and exploration well is not stationary at log-level. All variables are stationary at first difference and log first difference, so that the variables are integrated at a maximum order of one.

The ARDL framework is applied because it has the necessary flexibility to overcome the case of mixed order of integration. Table 2.11 reports the results from estimating ARDL model. It shows the empirical estimation of two single equations, with waiting time and discovery size as the dependent variables in each of the equations. The explanatory variables in each equation are the number of exploration wells and real oil price. It is essential to disaggregate waiting time and size because the high oil price affects the two differently. It is expected that a high oil price has a negative link with time between discoveries (shorter waiting time) and a positive link with discovery size. The optimal lag is chosen based on AIC, and it results in the model with ARDL(2,3,0) for the waiting time equation. Lag two is for waiting time, lag three is for number of exploration wells, and there is no lag for oil price. The ARDL(2,0,0) is applied for the discovery size equation, with lag two for size and no lag for both exploration well count and oil price.

Exploration wells and oil prices negatively affect the waiting time between discoveries. A 1% increase in the number of exploration wells from two years previously significantly reduces the waiting time between discoveries by 1.4 days. While, a 1% increase in real oil price shortens the waiting time by 0.5 days although this is not statistically significant. The second equation for discovery size shows that oil price has a positive relationship with the discovery size, whereas exploration wells have a negative relationship. A 1% increase in the number of exploration wells reduces the discovery size in the same year by 2.36%, while a 1% oil price increases the size by 0.57% but, again, this is not statistically significant in the short run.

Table 2.11 reports the bounds test result, adjustment factor, long-run equilibrium, and short-run dynamics coefficients. The first column of Table 2.11 presents the model estimated with the waiting time between discoveries as the dependent variable and the second column presents the log of discovery size as the dependent variable. The Pesaran et al. (2001) bounds test result indicates that a long-run equilibrium relationship is present among waiting time, the exploration well counts, and crude oil price. Panel (i) of Table 2.11 reports that the F-statistic and t-statistic of the adjustment factors are larger than the bounds test critical values at a 5% significance level. The null hypothesis that the level relationship does not exist is rejected for both equations; hence, there is enough evidence to support the long-run relationship of waiting time, exploration well counts, and real oil price at the level and so does discovery size well, and real oil price.

In Panel (ii) of Table 2.12, The adjustment factor ($-\alpha$) shows how the dependent variable changes when the three variables deviate from their long-run equilibrium. The adjustment factors for the two equations are all negative and statistically significant, which means that the estimated error correction forms represent stable relationships. The dependent variable of the prior period is too high relative to the long-run equilibrium, so it is necessary to decrease its value in the current period to revert to equilibrium. The adjustment factors of 0.786 and 0.734 suggest a rapid adjustment of the last year's waiting time and discovery size deviation, respectively, from the equilibrium.

In Panel (iii) of Table 2.12, the long-run equilibrium relationship is presented for the two equations. In the first column of Panel (iii), the long-run equilibrium for waiting time indicates that the coefficient is negative and significant for the exploration well counts and oil price shocks. A 1% increase in oil price relates negatively with the waiting time between discoveries by 0.7 days, while a 1% increase in well counts relates negatively with the waiting time by 0.9 days. The second column of Panel (iii) shows the long-run equilibrium for discovery size. The coefficient is positive for crude oil prices and negative for exploration well counts. A 1% increase in real oil price relates positively to the discovery size by 0.77%, and a 1% increase in well counts relates negatively to the discovery size by 3.2%. Panel (iv) shows that in the short-run relationship, waiting time is significantly affected by exploration well counts in the past one and two years, while oil price does not significantly affect waiting time between discoveries in the short run. Exploration well in the current year significantly affects discovery size, while oil price does not significantly affect discovery size in the short run.

Consistent with the literature, the crude oil price has a positive relationship with discovery size. Mohn (2008) argues that high oil price triggers exploration activity in the frontier area, with a prospecting aim of finding a larger size discovery than is available in the mature province. In the long run, a 1% increase in oil price is associated with a discovery size that is 0.77% larger, and a 1% increase in exploration wells is associated with a reduction in size of 3.2%. A rapid adjustment of 73.4% is necessary for discovery size in the current period to correct the deviation from the previous period. Similar to the case of waiting time, crude oil price also does not significantly affect discovery size in the short term, and exploration wells significantly affect the discovery size in the same period. A 1% increase in the number of exploration wells decreases discovery size by 2.36%.

2.6.2.1 The empirical ARDL model with a structural break

Table 2.13 shows the rejection of the null hypothesis of no structural break in supremum and exponential Wald tests, indicating the structural break's present. The tests result in the estimated break year in 1981 for the waiting time equation and 1999 for the discovery size equation. The negative shock of crude oil prices during the Global Financial Crisis 2008 can also affect the examined data series. Hence, three estimated break years are estimated in the ARDL model and presented in Table 2.14. Table 2.14a applies a dummy variable value of 0 from 1968 to 1980 and a value of 1 from 1981 to 2019. The coefficient estimates (γ_0) of the break year are statistically significant for both waiting time and discovery equations. One of the strong reasons that the break occurred in 1981 is because there were first discoveries, Asgard in the Norwegian Sea, which became an important Norway's petroleum industry, following a big accident near the Ekofisk area in 1980. The structural break in 1999 follows the large NCS discovery size of Ormen Lange, and the break year of 2008 represents Global Financial Crisis. Table 2.14b presents the ARDL estimates with an estimated break year in 1999. Table 2.14b presents the ARDL estimates with an estimated break year in 1999. A dummy variable value of 1 is applied for 1999 onwards, and a value of 0 from 1968 to 1998. For this case, the coefficient estimate of the break is only statistically significant for discovery size, as indicated earlier in the supremum wald test. A similar result is obtained when the estimated break year is in 2008, as shown in Table 2.14c, in which the break in discovery size is significant.

Table 2.13: Structural break test result for waiting time between discovery, discovery size, exploration well, and crude oil price

Equation	Supremum wald	Exponential wald	Estimated break year
wt	43.545***	18.668***	1981
s	13.351	5.411*	1999

Notes:
Standard error in parentheses; * $p < .10$, ** $p < .05$, *** $p < .01$
The critical value of Wald test is based on Hansen (1997)
and the critical value of recursive test
is based on Ploberger and Krämer (1992).

Among the model with estimated break years presented in Tables 2.14a-2.14c, the relationship between waiting time between discoveries, well counts, and oil price is consistent in terms of the sign and statistical significance. The real oil price in the current year and the exploration well from the previous year are statistically significant on the waiting time between discoveries. The high oil price and more oil exploration well count reduce the time between findings. Oil exploration companies have the motivation to discover more oil and increase the exploratory effort indicated by more exploration well, which leads to a shorter time between oil discoveries. The three estimated break years also confirm the positive relationship between crude oil price and discovery size. Real oil prices in the current year significantly affect the average oil discovery size. Based on the break year in 1999 for discovery size, the exploration well significantly affects the size with three years lag. In terms of the relationship between well and size, after including a break, a positive relationship is found.

Table 2.14: ARDL estimation with a structural break for exploratory effort, efficiency, and crude oil price

(a) Estimated break year in 1981

	(i) wt_t	(ii) ls_t
Coefficient estimates ($\phi_i, \beta^{'}_{j,i}, \gamma_0$)		
wt_{t-1}	-0.104 (0.133)	
wt_{t-2}		
ls_{t-1}		0.291** (0.139)
lw_t	-56.797 (52.733)	0.086 (0.641)
lw_{t-1}	-189.139*** (56.460)	
lop_t	-80.565** (32.715)	0.382 (0.467)
b_t	156.016*** (43.954)	-1.321** (0.633)
(a_0)		
Constant	1255.100*** (195.981)	0.973 (1.962)
Observation	48	53
F-statistic	10.13	4.22
Prob (F-statistic)	0.0000	0.0056

Standard Error in parentheses; * $p < .10$, ** $p < .05$, *** $p < .01$
The notation "l" indicates natural logarithmic of the variable,
e.g. lw_t indicates logarithmic value of exploration wells at period t.

Offshore Norge (2017) states that the peak oil production in Norway was in 2000, with a new government scheme introduced to stimulate exploration. That could be one of the reasons for an overall positive relationship between well and size after the break in short run, with more findings in new fields.

The consistency in the long-run equilibrium relationship is presented in Table 2.15 for the ARDL model with a break year in 2008. Adding a structural break in the model confirms the consistency of both equations' long-run equilibrium relationship. There is a long-run equilibrium relationship between the waiting time, oil exploration well counts, and oil price, as well as between discovery size, well, and oil price, as indicated by a rejection of Bounds test in Panel (i). The long-run equilibrium is found to be stronger in discovery size than the waiting time equation. Panel (ii) shows the sign of adjustment factors are negative and statistically significant for both equations. It is indicated that there is a rapid adjustment by 74.5% at current year when

(b) Estimated break year in 1999

	(i) wt_t	(ii) ls_t
Coefficient estimates ($\phi_i, \beta'_{j,i}, \gamma_0$)		
wt_{t-1}	-0.035 (0.146)	
ls_{t-1}		0.229* (0.122)
ls_{t-2}		-0.341*** (0.122)
ls_{t-3}		0.092 (0.123)
ls_{t-4}		-0.393*** (0.119)
lw_t	-14.616 (57.454)	-1.103 (0.687)
lw_{t-1}	-130.642** (58.886)	1.506* (0.818)
lw_{t-2}		-1.481* (0.788)
lw_{t-3}		1.446** (0.586)
lop_t	-120.157*** (38.247)	2.823*** (0.726)
lop_{t-1}		-1.440* (0.823)
b_t	52.357 (32.944)	-2.624*** (0.536)
(a_0) Constant	1152.838*** (218.481)	-2.053 (1.831)
Observation	51	53
F-statistic	6.71	5.08
Prob (F-statistic)	0.0001	0.0001

Standard Error in parentheses; * $p < .10$, ** $p < .05$, *** $p < .01$
The notation "l" indicates natural logarithmic of the variable,
e.g. lw_t indicates logarithmic value of exploration wells at period t.

the waiting time deviates from the equilibrium in the previous year and by 69.9% adjustment for discovery size. Panel (iii) shows that the sign of coefficient estimates in the long-run relationship is consistent with the model without a structural break, except for well and size relationship. A 1% increase in crude oil price relates negatively with the waiting time by a day and relates positively with discovery size by 1%. A 1% increase in the exploration well counts relates negatively to the waiting time by a day and relates positively with discovery size by 1.36%. The negative relationship between well and discovery size is found in the short run. As indicated in Panel (iv), the break is statistically significant for the discovery size equation, which causes a reduction in the size.

As a robustness test, the nominal oil price is adjusted with Purchasing Power Parities (PPP) to overcome the exchange rate fluctuations for non-

(c) Estimated break year in 2008

	(i) wt_t	(ii) ls_t
Coefficient estimates $(\phi_i, \beta'_{j,i}, \gamma_0)$		
wt_{t-1}	0.013 (0.149)	
wt_{t-2}	0.242* (0.141)	
ls_{t-1}		0.301** (0.143)
lw_t	-28.747 (59.056)	-0.236 (0.766)
lw_{t-1}	-141.342** (64.387)	1.188 (0.759)
lw_{t-2}	-19.555 (68.136)	
lw_{t-3}	113.744** (52.481)	
lop_t	-71.647* (41.010)	2.120** (0.834)
lop_{t-1}		-1.393 (0.890)
b_t	-6.727 (48.596)	-1.491** (0.659)
(a_0)		
Constant	693.220** (312.015)	-4.118 (2.888)
Observation	48	49
F-statistic	4.73	3.37
Prob (F-statistic)	0.0004	0.0084

Standard Error in parentheses; * $p < .10$, ** $p < .05$, *** $p < .01$
The notation "l" indicates natural logarithmic of the variable,
e.g. lw_t indicates logarithmic value of exploration wells at period t.

US fluctuations. The detailed result of the model is presented in Table A.1, Appendix A.2. It shows that the long-run equilibrium relationship is present among waiting time, exploration well counts, and real oil price, and so does discovery size, exploration well counts, and real oil price. The adjustment factor is negative and statistically significant, indicating a stable equilibrium relationship. The sign coefficient estimates are also consistent with the model applying U.S. CPI as a deflator of nominal oil price. In the long run, real oil price relates negatively to the waiting time between discoveries and positively to discovery size. Exploration well counts relate negatively to the waiting time and discovery size in the long run.

The simulation and empirical results align with the economic intuition that high oil price boosts exploratory effort. As a consequence, the waiting time is shorter, with many exploration companies being incentivised to explore.

Table 2.15: ARDL estimation in error correction form with a structural break in 2008 for exploratory effort, efficiency, and crude oil price

	(i) $D.wt_t$	(ii) $D.ls_t$
(i) *Case 3*		
Bound Test H_0: no level relationship		
F-stat‡	4.821^{*}	8.113^{***}
t-stat‡	-3.764^{**}	-4.897^{***}
(ii) *Adjustment factor (-α)‡*		
wt_{t-1}	-0.745^{***} (0.198)	
ls_{t-1}		-0.699^{***} (0.196)
(iii) *Long-run* (θ)		
lw_{t-1}	-101.875 (74.557)	1.361 (1.145)
lop_{t-1}	-96.167^{*} (49.397)	1.040 (0.727)
(iv) *short-run* ($\psi_{yi}, \omega', \psi'_{xi}, \gamma_0$)		
$D.wt_{t-1}$	-0.242^{*} (0.140)	
$D.ls_{t-1}$		
$D.lw_t$	-28.747 (59.056)	-0.236 (0.766)
$D.lw_{t-1}$	-94.189 (65.237)	
$D.lw_{t-2}$	-113.744^{**} (52.481)	
$D.lop_t$	-71.647^{*} (41.010)	2.120^{**} (0.834)
b_t	-6.727 (48.596)	-1.491^{**} (0.659)
(a_0) Constant	693.220^{**} (312.015)	-4.118 (2.888)
Observations	48	49

Notes:
Standard error in parentheses; * $p < .10$, ** $p < .05$, *** $p < .01$
‡The approximate p-values applied for speed-of-adjustment coefficient is based on Kripfganz et al. (2018)

The negative relationship between the discovery size and the exploration wells must be interpreted more carefully. Exploration well count is associated with a decline in size because the higher frequency of the lower average discovery size is associated with the result of more exploratory effort. The giant discovery size is mostly discovered during the early life cycle of the basin, then, as more discoveries are made (more exploratory effort), the discovery size is smaller. However, as suggested by Mohn (2008), this must be interpreted carefully when taking into account exploration risk. Mohn (2008) suggests that in response to high oil price, the risk appetite is high, and there is an increase in the exploratory effort in the frontiers area with high expected discovery size.

2.7 Conclusion

Chapter 2 attempts to answer two research questions: First, how does exploration activity behaviour change over time? Second, how does crude oil price fluctuation affect exploration? This study contributes to the existing academic literature in two ways. First, it emphasises the important role played by oil exploration in determining future oil supply in the long run. It does this by conducting a simulation exercise to uncover the effect of exploratory effort. Second, it includes the economic variable of oil price in the exploration model. The empirical ARDL framework is applied to understand how exploratory effort and crude oil price fluctuation affect exploratory efficiency in the long and short run.

To address the first research question, a Monte Carlo technique is applied to simulate two variables in a mature petroleum province (i.e., NCS): the time in between two discoveries are made and discovery size. The simulation exercise applies parameter changes in order to capture the changes in exploratory effort, which are illustrated by the density shape and matrix statistical analysis. A key finding in the simulation of waiting time is that the more frequent the exploratory effort, the shorter the waiting time between discoveries and the shorter the time required to reach the last discovery. The simulation of discovery size finds that it is more frequent to discover small discovery sizes than large ones. Combining these outcomes implies that a shorter waiting time between discoveries leads to more findings of the smaller discovery size. The smaller discovery size is found after the giant one, and the discoveries are more frequent. The results are consistent with economic intuition and the existing literature, which assert that giant discovery is mainly found in the earlier basin life-cycle, and is less frequently found than the small discovery.

As regards the second research question (i.e., how crude oil price fluctuation affects exploration), the ARDL empirical model is applied to analyse the relationship between crude oil price, exploratory effort, and efficiency. The empirical estimates conclude that a long-run relationship is present between crude oil price, exploratory effort, and efficiency. Crude oil price affects exploratory efficiency in the long run but not in the short run. In contrast, exploratory effort significantly affects efficiency in the short run. The empirical analysis offer a clearer interpretation of how oil prices and exploratory effort affect efficiency when the waiting time and size estimates are distinguished. The relationship between discovery waiting time and oil price must be interpreted carefully. In the long run, oil price and the number of exploration well relate negatively to the time in between discoveries. The short-run dynamics also show a negative relationship between crude oil price and waiting time. The crude oil price negatively affects the waiting time because a high oil price triggers oil producers to drill more. As exploratory effort increases frequency, the number of exploration wells increases, and this is associated with a reduced waiting time between discoveries. The effect of exploratory effort on waiting time between discoveries can be captured significantly over three years. The empirical result suggests that oil price relates positively to the discovery size in the long-run. The short-run dynamics support the existing literature that suggests there is a positive relationship between crude oil price and discovery size. A high oil price incentivises oil producers to discover potential frontier oil fields, which are of a relatively larger size than the existing fields. The more frequent exploratory effort is associated with a smaller average discovery size, which explains the negative relationship between exploration well and size.

This study develops a toolbox for Monte Carlo simulation that can be adjusted to data from a different oil region. For instance, in the simulation result of NCS, there is a very high probability that the average waiting time between oil discoveries is less than ten months, with the average discovery size being smaller than the giant discovery. A similar exercise can be applied to the other petroleum provinces to comprehend the range of the parameter changes. The lesson learnt from this case study of a mature petroleum province is also helpful to initial studies looking at the exploration plan in the frontier fields with higher risk. For instance, the simulation exercise illustrates the probability of a specifically-sized well being discovered within certain periods. Applying the same framework, the parameters of Monte Carlo simulation can be approached by using the estimates of what has most likely occurred in order to establish the confidence level of certain discovery sizes and the lag time between such discoveries in the frontier fields, irrespective that there is less historical data. The empirical result acknowledges a long-run relationship between exploratory effort, efficiency, and oil price. Together, these affect how future oil supply responds to crude oil price shocks. Understanding this will assist decision makers to design policy measures that boost exploration activity in response to extreme oil price movements.

Chapter 3

World oil production, global demand, and the crude oil price: evidence of the asymmetric effect

3.1 Introduction

The role of supply and demand has been extensively discussed in the current literature on the global oil market. Global demand is associated with global economic activity, while crude oil supply is a result of current production and past exploration activity. Most of the recent literature asserts that demand plays a more significant role in crude oil price shock than supply even though, historically, oil supply disruption has had significant impact on crude oil price fluctuation. Previous studies by Kilian (2009); Kilian and Murphy (2014) strongly assume that oil supply causes only small and temporary shocks in the crude oil price increase whereas aggregate demand plays an important role in causing large and persistent shocks to the rise of oil prices. However Baumeister and Hamilton (2019) propose a weakening of this assumption in a re-examination of the roles played by supply and demand. These authors conclude that the supply-side plays an essential role during historical oil price movement. They note that the effect of strong supply was important to the oil price collapse in 2014–16 and that strong demand led to the oil price bounce-back later in 2016. Kolodzeij and Kaufmann (2014) evaluate the robustness of Kilian's (2009) conclusion about the unimportance of oil supply by disaggregating OPEC and non-OPEC production, and they find a negative long-run relation between OPEC and non-OPEC production and oil price. Ratti and Vespignani (2015) argue that OPEC production is more responsive to oil price shock in the new industrial age (i.e., from 1997 to 2012) compared to the 1974 to 1996 era, which is responsive to non-OPEC production.

Previous studies have several shortcomings in their modelling of the relationship between crude oil price, supply, and demand. First, the crude oil price is assumed to affect global oil supply and demand symmetrically, which ignores the different effects of oil price increases and oil price decreases. It is necessary to take this asymmetric effect into account when advising oil producers how to overcome an extreme rise and drop in oil prices by anticipating and adjusting oil production. The extant literature that does consider the asymmetric effect on oil prices is mainly concerned with the economic activity that influences global demand, with scant attention being given to discussing the asymmetric effect of crude oil price on global oil supply. The literature suggests, for example, that a rise in crude oil price is associated with a slow business cycle (Hamilton, 1983, 2003; Sadorsky, 1999; Apergis et al., 2015). Mork et al. (1994) find evidence of an asymmetric relationship between the increase and decrease in oil prices and GDP growth in the industrialised countries between 1967 and 1992, particularly for the US, Japan, and Norway. Sadorsky (1999) concludes that oil price volatility shocks have asymmetric effects on industrial production and real stock returns. Hamilton (2003) argues that oil price increases are more significant that oil price decreases in predicting GDP growth. Lardic and Mignon (2008) also support the earlier literature in finding that an increase in oil price is more damaging to economic activity than a decrease. Apergis et al. (2015) find evidence of an asymmetric long-run relationship between oil price and US economic growth.

Second, most current literature only analyses the relationship between oil price, supply, and demand in the short run, and there is a strong assumption that production does not respond to oil price shocks. This is true to an extent, given that production is not flexible enough to make adjustments in the short term, but it is essential to take the long-run relationship into account. Further, there is a lag of a few years between the completion of an oil exploration stage and actual production. Analysing this by applying a structural VAR model requires the variables to be treated in the first difference, eliminating the long-run relationship. Only a few studies take into account the long-run relationship between oil price, supply, and demand, including Kolodzeij and Kaufmann (2014) who examine the long-run relationship between OPEC production, non-OPEC production, the real economic activity index, and oil price. There is therefore a gap in the literature concerning how the global oil supply, represented by world oil production, responds to oil price movements.

Third, in most of the literature, the stationarity property of the variables is tested using the standard unit root test. Furthermore, a structural break is rarely incorporated in the empirical model, while it is essential given that the time series, remarkably crude oil price and global demand, contains several extreme fluctuations. This leads to a bias when there are breaks in the time series. Applying a unit root test that accommodates structural breaks and taking into account an estimated break date in the empirical model can address this. Such a test is also necessary to determine which empirical model is suitable for analysing the relationship among the variables. In the case that the variables are mixed stationary and non-stationary, neither structural VAR nor cointegrated VAR model can be applied. This is because structural VAR requires all variables to be stationary, and cointegrated VAR requires all variables to be integrated in the same order.

Therefore, there are clearly shortcomings in the current literature. This chapter attempts to address these by answering two research questions: what is the relationship between crude oil supply, demand, and oil prices? How do positive and negative shocks in the crude oil price affect oil supply and demand differently? This study is interested in identifying whether any long-run equilibrium exists between world oil production, global demand, and crude oil price. The long-run relationship could have policy implications for oil producers' decision-making, such as whether to invest in more extensive capital and technological change, whether to adjust the volume of the oil produced, stored, and exported, and how to assess the investment strategy with regard to the expected growth in oil price. Secondly, this study investigates whether asymmetric effects matter in the global oil market. The importance of accounting for an asymmetric effect will improve understanding of which crude oil price changes (i.e., spikes versus drops) have a more noticeable impact on the global oil supply and demand. World oil production and economic activity may react differently to positive and negative shocks in the crude oil prices.

The main differences between this study and the existing literature, most notably Kilian (2009)'s work, are threefold. First, there is the treatment of the time series variables used; second, it uses a unit root test that accommodates the possible breaks and takes into account a structural break in the empirical analysis; and third, it identifies the asymmetric effect that decomposes the shocks into partial positive and negative changes.

In more detail, this study first provides an alternative empirical solution to the structural VAR model proposed by Kilian (2009) and Kilian and Murphy (2014) by examining the global oil market variables in levels so that the long-run relationship is taken into account. The treatment of the variables in levels contrasts with Kilian (2009), whose analysis uses first differences. Variables in the first difference eliminate the long-run relationship between world oil production, global demand, and oil price, creating a tendency to underestimate the supply-side's role. As one of the objectives is to analyse the long-run equilibrium relationship, applying variables in levels is preferred to first differences so as to avoid losing the long-run relationship.

Second, this study checks the structural breaks in the temporal dynamics of each time series and the relationships across the three series. A unit root test that considers the possibility of the structural break in each series is utilised to determine the suitable econometric framework. This chapter applies the Enders and Lee (2012) unit root test, which uses a Fourier transformation to identify possible breaks. The benefits of applying the Enders and Lee (2012) unit root test are as follows. First, there is no need to specify the number of the breaks nor the time when they occur, and second, it is sufficiently flexible to be able to cope with both sharp or smooth breaks. The Enders and Lee (2012) unit root test results indicate that the variables are in a mixed order of integration, with oil production and global demand being I(0), and crude oil price being I(1). In this case, the modelling needs to accommodate the mixed order of integration among the underlying variables in levels. Hence, the use of the ARDL framework, in which this study differs from Kolodzeij and Kaufmann (2014). The ARDL framework has the flexibility to deal with the mixed non-stationary and stationary issues. ARDL can obtain consistent estimates for long-run equilibrium whether the underlying variable is stationary or not, as long as it is not integrated at order 2: I(2) (Pesaran and Shin, 1997). The other advantage of applying the ARDL framework is that it can disentangle the short-run and long-run equilibrium relationship through its error correction form.

A structural break is incorporated in the ARDL model, and the estimated break date is January 2009, following the Global Financial Crisis. The break is found to be strong in global demand and crude oil prices, which causes a reduction in both demand and oil prices. Crude oil production is not affected by the break as the oil production cannot be adjusted quickly and it may take longer to respond to the economic changes. Adding a model with a break also confirms the presence of the long-run equilibrium relationship between supply, demand, and oil price. The ARDL model with a break also has a consistent sign in the coefficient estimates compared with the model without a break.

Third, the asymmetric effect is taken into account in the empirical model. The variables are decomposed into partial positive and negative changes, and asymmetry represents the difference between a positive and negative change in the variables. The asymmetric effect is modelled by non-linear ARDL (NARDL) proposed by Shin et al. (2013), distinguishing the long-run equilibrium from the short-run dynamics relationship. The monthly time series of world oil production, the Baltic Dry Index, and the real oil price from January 1985 to December 2019 are estimated in the basic ARDL and asymmetric ARDL (NARDL). Both the basic and asymmetric ARDL models support the presence of a long-run relationship between world oil production, global demand, and crude oil price, except that the oil price response to asymmetrical oil production and demand is weak.

Adding a structural break in the NARDL model also confirms that the break is significant in global demand and crude oil prices. The interaction of the break with the positive and negative global demand shocks strongly affects the crude oil price, while the break interaction with the negative shock of oil supply on oil prices is stronger than with the positive shock. The interaction of the positive and negative supply shocks significantly affect global demand. Consistent with the model without a break, supply does not respond to the break in the short run. NARDL model with a break in January 2009, also confirms that the long-run equilibrium relationship exists. In the NARDL with a break, the positive and negative shocks are significant on the crude oil prices in the long run, while supply shocks are not. The positive supply and price shocks are significant in global demand.

The remainder of the sections is organised as follows. Section 3.2 discusses the existing literature on crude oil production, global economy, and real oil price. Section 3.3 describes the data and variables applied in this chapter. The econometric framework of basic ARDL and NARDL is described in Section 3.4. Section 3.5 presents the empirical results, and Section 3.6 concludes the chapter.

3.2 The literature on crude oil price and production and contribution of the study

Many studies have investigated the cause of crude oil price fluctuation since crude oil become a global market in the early 1980s. They have been applying different models that are mainly employing the VAR approach. In past studies, refer to Hamilton (2003) and Kilian (2008), crude oil supply shocks are widely believed to play an essential role in real oil prices. This view has become an open debate since the global oil market developed then this notion has been reinvestigated in the current research.

The analysis of more recent oil price fluctuation has been brought into the existing literature by Kilian (2009) seminal paper that identifies the effects of underlying demand and supply shocks in the global crude oil market between 1973 and 2009. Kilian (2009) analyses the effects of the oil supply and demand on the oil price by applying structural VAR. In that seminal paper, three main structural shocks have been employed: flow supply, aggregate demand, and precautionary demand shocks. Kilian (2009) argues that the effects of the real oil price vary depending on the shocks to drive the oil price fluctuation and emphasises the importance of aggregate demand and precautionary demand that play an essential role and contribute more to driving historical oil price fluctuation. Flow supply shock represents a disruption in the global oil supply proxied by changes in crude oil production. Flow demand is associated with the boom in the global business cycle that causes an increase in crude oil demand as the primary energy. Kilian (2009) constructs the real economic activity index and utilises the index to represent global demand. The precautionary demand refers to the forward-looking behaviour of oil above the ground. Kilian (2009) applies the real acquisition of the U.S. imported oil price to proxy the fluctuation in speculative demand. This type of shock is caused by the expectation of an oil supply shortage relative to oil demand in the future.

3.2.1 The extension of global crude oil market model

Many pieces of literature extend Kilian (2009) seminal paper to explain the crude oil price fluctuation. The first extension takes into account the speculation role that has been extended by Kilian and Murphy (2014). The second extension disaggregates the production data into several categories: OPEC and non-OPEC, crude oil exporter and importer countries, and conventional and unconventional oil production. The third extension considers the reduced-form VAR model as the extended work of structural VAR in Kilian (2009) to compare the price elasticity between those models and create the oil price forecast. Another extension includes incorporating other factors contributing to the crude oil price shocks by varying the sample period, data sets, data treatment, and methodology that lead to different results.

3.2.1.1 The role of speculation

The role of speculation has been taken into account in Kilian and Murphy (2014) to extend Kilian (2009). Crude oil stock as the proxy of the speculation reflects the market expectation of the availability of the future oil supply. The market expectation also includes anticipating future events that trigger the shift in the crude oil inventories above the ground.

Based on the economic theory, Kilian and Murphy (2014) extends the previous Kilian (2009)'s work to understand better the dynamic effects of the shocks on real oil price by restricting the sign of the response of global crude oil market variables to each shock. The negative flow supply shocks shift the supply curve to the left, causing a rise in crude oil prices and drops in global demand. The positive flow demand shocks cause the demand curve to shift to the right, implying the increase in real oil price and triggers in crude oil production increase. A positive speculative demand shock is associated with the jump in inventories that expect the high oil price as the news about future supply shortages. This shock is associated with market concern that shifts oil prices. Kilian and Murphy (2014) do not put any sign restriction on the inventories as the effect is still unclear. Oil stocks decrease as global demand drops. However, inventories can also rise to anticipate the jump in oil prices. By incorporating speculative demand shock in the global oil market model, the supply shocks contribute to the oil price shocks compared to the previous result in Kilian (2009). Then, they conclude that although speculation did not play an essential role in the oil price fluctuation between 2003 and 2008, there were some episodes in the past, namely after the Iranian revolution in 1979, after the OPEC collapse in 1986, and after the Iraqi invasion of Kuwait in 1990 that speculative shocks affect the oil price decline.

3.2.1.2 The use of disaggregated data

The second extension includes disaggregating the production data based on producer characteristics, namely OPEC and non-OPEC, exporter and importer countries, and conventional and unconventional oil production. OPEC and non-OPEC have different methods of setting their production criteria and quota that trigger them to respond differently to the crude oil price shock. Ratti and Vespignani (2015) argues that OPEC production exhibited a positive and significant response to the crude oil price shock from 1997 to 2012 but no significant response in the previous era, from 1974 to 1996. A 1% rise in the real oil price increases the non-OPEC oil production by 0.05% over 1997-2012. Non-OPEC production showed a positive and significant response to the crude oil price from 1974 to 1996 but no considerable response from 1997 to 2012. A 1% rise in the real oil price increases the non-OPEC oil production by 0.02% from 1974-1996.

Cologni and Manera (2014) identify a positive and significant relationship between oil price fluctuation and some of the OPEC countries, namely Saudi Arabia, Algeria, and Kuwait, and no significant relationship between oil price and most of the non-OPEC countries except Norway, Mexico, and Canada from January 1995 to December 2009. In the smaller quarterly sample size between 1984 and 2002, Dees et al. (2008) also conclude that non-OPEC production has little response to the real oil price changes with a 1% rise in real oil price increasing oil production 0.02% for most non-OPEC countries. Mexico and the U.S exhibit elastic supply, with their production increases by 5.5% in response to a 1% rise in real oil price. They argue that this high elasticity is driven by capability of a large number of producers to increase production in the U.S. and by the incentive to grow the revenues in Mexico. By disaggregating the production data into crude oil exporter and importer countries, Vu and Nakata (2018) find that the response to the global oil market differs between oil exporter and importer countries.

The most recent extension of the world oil market model disaggregates production into conventional and unconventional oil production. Conventional and unconventional crude oil production have different technology and methodology to extract crude oil due to their distinct characteristic. Many studies investigate the response of unconventional oil production to crude oil price shocks focusing on the U.S. scale since the U.S. is the biggest contributor to the world's unconventional oil production. Unconventional oil production, particularly tight oil, has been developed in the U.S. as the most significant global unconventional oil producer since 2000, followed by unconventional Canadian oil production such as oil sands, tight and shale oil, and bitumen since 2005. Tight oil accounted for approximately 61% of the total U.S. crude oil production in 2018 (EIA, 2019) and 8% of total world crude oil production. Tight oil, shale oil, and oil sands accounted for 65% of Canadian oil production and 1.5% of the world's crude oil production in 2016 (NEB, 2018).

Kilian (2017) argues that the fracking boom in the U.S is not the main reason for the oil price decline between 2014 and 2015. Prest (2018) extends Kilian (2017) work by using U.S. crude oil production, world oil production, and oil rig count series and supports the view of Kilian (2017) that little evidence on the effects of U.S shale on the oil price drop during 2014-2016, despite the weak global economic activities that drive the oil price decline in that period.

3.2.1.3 Reduced-form VAR implementation

Reduced form VAR can be applied to forecast and examine the magnitude of the shocks contributing to the oil price decline in a certain period. Baumeister and Kilian (2016*b*) use reduced-form VAR to investigate the cause of the oil price decline in the episode between June and December 2014. The model can predict oil prices and highlight the cause of the oil price decline caused by negative demand shock, positive supply shock, unpredictable shocks due to oil price expectations, and the unexpected global economic slowdown. Applying structural and reduced-form VAR models will give different magnitudes on the price elasticity impact.

3.2.1.4 The difference in the time series data treatment

Various research treats the time series data differently, such as focusing on the specific length of the sample period, applying the series in level or first difference, and considering the structural break over the period. The well-established current studies commonly examine the production response within the short to medium sample period and treat the data as non-stationary; $I(1)$ in level. Kilian (2009) transforms the series into the first difference to be stationary and apply VAR in the first difference, while cointegrated VAR in level is applied in Kolodzeij and Kaufmann (2014) work. Implementing the VAR in the first difference eliminates the adjustment term and cointegrating vectors of the long-run equilibrium. Only the short-run dynamics relationship can be examined using this model. The cointegrated VAR model can identify the long-run equilibrium relationship among the variables. However, cointegrated VAR cannot be applied in the mixed order of integration among the variables, e.g., $I(1)$ and $I(0)$.

In the time series, there is a circumstance called the structural break in which the data sets' fluctuation over time may exhibit mean, trend, or slope shifts. The consequences of those shifts will lead to spurious regression, and the model will be misspecified. The inclusion of the structural break has been well described in the current literature, such as in Fan and Xu (2011) and Banerjee et al. (2017), with the common finding that by considering a structural break, the driving force of the oil price fluctuation and the relationship among the variables within the particular period can be more precisely identified. The results may significantly differ before and after the structural break. Dividing the entire sample into several subsamples gives a more straightforward interpretation than the general conclusion obtained from the total sample without considering the break. The application of structural break can be in the unit root test for each series and the relationship of the variables in the long-run equilibrium.

Fan and Xu (2011) incorporate structural break in the unit root test for crude oil data series and argue that structural break is essential to understand the main drivers of the oil price fluctuation in the short-run and long-run time horizon. Their work analyses the crude oil market using data from January 2000 to September 2009 and finds two breaks in March 2004 and June 2008. The break test is based on the moving-estimate test by Chu et al. (1995). The structural break test by Lee and Strazicich (2003) is also applied and obtains the breaks in September 2004 and February 2005 for the level shifts or September 2001 and July 2008 for the level and slope trend changes. They argue that the driving factors across the oil price fluctuation are significantly distinct in the different periods. Before the first break; between January 2000 and March 2004, the main drivers were caused by various short-term factors such as speculation and geopolitical events, between the first break and the second break; from March 2004 to June 2008, they were caused by financial market and speculation, and after the second break-in June 2008, they were caused by supply and demand changes. Those short-term factors interact with each other and form the supply and demand fundamentals in the long term. Without considering the structural break, all of the factors seem to be significant. It fails to show the fundamentals in each episode with the multicollinearity possibility and lower explanatory power that leads to less accurate interpretation.

Other studies applying structural break for modelling oil price volatility are by Salisu and Fasanya (2013). Their study applies structural break by Narayan and Popp (2010) in the unit root test and then asymmetric GARCH to model the oil volatility. As the work examines crude oil price characteristics, the model is univariate series, and other factors are not considered. Furthermore, the data treatment also employs crude oil price in growth rate instead of level. The current literature's limitation focusing only on break in stationarity test motivates this chapter to investigate the structural break in the regression model with multivariate series.

3.2.1.5 The inclusion of other factors in the modelling

Different proxies lead to different outcomes; for instance, applying other factors affecting crude oil price as the independent variables, the various proxies for global demand, and crude oil price fluctuation analysis at the country level. Other factors that may affect the rise and drop of crude oil prices are also investigated with different modelling approaches. These factors include geopolitical events, financial market conditions, OPEC capacity and behaviour, and downstream factors. Kaufmann et al. (2004); Dées et al. (2007); Dees et al. (2008) apply dynamic OLS, and Fan and Xu (2011) use a multifactor market model to estimate the relationship among those other factors.

Existing studies apply various indices to global demand, such as Kilian (2009) 's index, Baltic Dry Index (BDI), industrial production, real GDP, and commodity index as the demand proxies. Some existing literature also focuses on analysing the production response to the crude oil price at the country level.

Anderson et al. (2014) argue that oil production from the existing wells in Texas does not respond to the oil price fluctuation of spot and futures prices, and the geological constraints mainly cause this. They apply regression in the first difference and conclude that a one percentage point increase in the expected rate of the oil price increase causes the decline in oil production in Texas by 0.1% between 1990 and 2007. One percentage point increase in the oil price decreases the oil production by 0.4% when the spot price is applied. None of those estimated coefficients is statistically significant. Moreover, they support Kilian (2009) finding that unanticipated positive demand shock causes an immediate rise in oil price while production increases but not immediately. For a large shock, the gradual increase in production will cause a gradual fall in oil prices.

3.2.2 The contributions to the existing literature

Based on the current empirical findings discussed above, certain gaps have been found in the existing literature. First, the robustness of the unit root test was applied to conclude the stationarity of the variables. Then, it leads to the suitability of the modelling approach to understand the relationship among the variables. Second, the importance of structural break in the cointegrating vector has not been taken into account. Third, they do not allow the asymmetric response of oil production to the long-run equilibrium. Fourth, the structural break is rarely discussed in the empirical model to understand the relationship between supply, demand, and crude oil prices, while there have been many extreme oil price fluctuations.

The current studies most likely found that all series are integrated order one in a relatively short-term sample period, then apply VAR. However, VAR is less powerful since it cannot distinguish between short-run and long-run dynamic relationships. Furthermore, VAR is appropriate if all the series are stationary in level or $I(0)$. Another difference in modelling includes a standard cointegration approach that can identify the long-run relationship among the variables. However, the standard cointegration model also has the shortcoming that all series are supposed to be integrated in the same order. Hence, these models do not appropriate when there is evidence of a mixed order of integration; $I(0)$ and $I(1)$. Alternatively, the ARDL model allows the data series to be integrated in the mix order as long as they are not $I(2)$. Moreover, apriori-knowledge of the sign restrictions of the variable is not required in the ARDL model.

This research investigates three objectives; first, this chapter contributes to the methodological application of the structural break in the global oil market. The stationarity of the time series incorporating the structural break is examined using different testing methods, notably, the Enders and Lee (2012) test that does not require determining the break dates or the number of the breaks. A structural break is taken into account not only in the temporal dynamics of each time series but also in the relationships across the world oil production, global demand, and crude oil prices. A supremum Wald test is applied to estimate the break date to be incorporated in the ARDL model. Third, this study examines the asymmetric response of crude oil production to the crude oil price deviation from long-run equilibrium. Lastly, the production response to the deviation from long-run equilibrium is analysed with the structural break.

3.3 Global oil market data and descriptive analysis

3.3.1 Description of the data

This section describes the crude oil supply, demand, and real crude oil price data examined in the empirical analysis. The datasets are time series from January 1985 to December 2019, consisting of monthly data for the (logged) level of world crude oil production (prod) as a proxy for world oil supply, the (logged) level of Baltic dry index (BDI) as a proxy for global demand, and the (logged) level of the real acquisition cost of imported oil price (op) as a proxy for the crude oil price.

Monthly observations of world crude oil production are used to measure global crude oil supply; these are obtained from the U.S. Energy Information Administration (EIA).[1] The supply proxy is used to indicate concern about the physical availability of crude oil. The world crude oil production is measured in Million barrels per day (Mb/d).

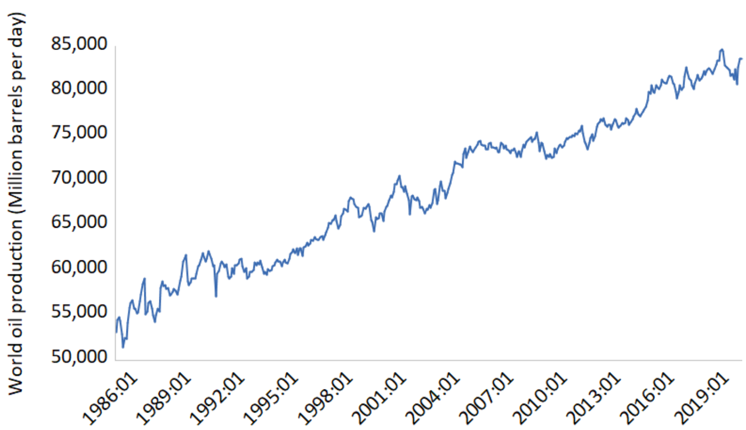


Figure 3.1: World crude oil production

[1]https://www.eia.gov/international/data/world

The global demand is proxied by the Baltic dry index (BDI), a composite of the dry bulk shipping market index; this measures the fluctuation in the global business cycle. The monthly series is available on the Baltic exchange website.[2] The index considers the freight rates for carrying various bulk dry commodities across 31 routes worldwide. BDI is measured as a cost, so the nominal value is deflated by Consumer Price Index (CPI) for all urban consumers, thus taking into account the inflation rate throughout the years. The underlying motivation for using BDI is that the freight rate increases when the demand for cargo increases, and the demand for cargo increases when the global business cycle increases. The CPI series are obtained from the U.S. Bureau of Labor Statistics, and they are also available from the U.S. EIA.[3]

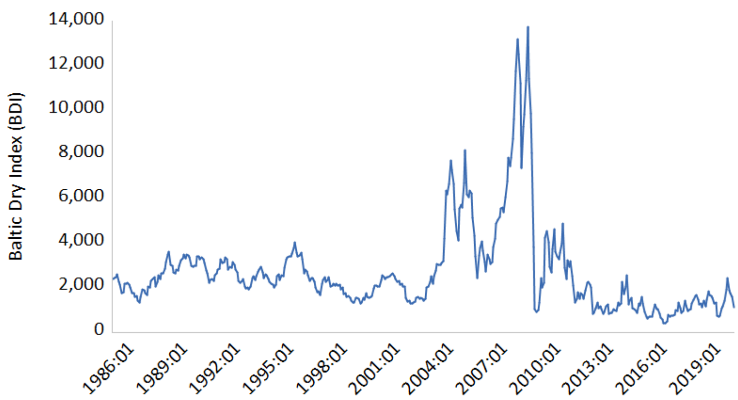


Figure 3.2: Baltic Dry Index

The crude oil price is proxied by the U.S. refiner acquisition cost of imported crude oil in USD per barrel; this is obtained from the U.S. EIA.[4] The real price is preferred to the nominal price because the real price takes into account the inflation rate, which is more useful for long-term fluctuation analysis. Thus, the nominal crude oil price is deflated by Consumer Price Index (CPI) for all urban consumers to generate the real price.

[2]https://www.balticexchange.com/en/data-services/market-information0/dry-services.html
[3]https://www.bls.gov/cpi/data.htm
[4]https://www.eia.gov/outlooks/steo/realprices/

The data employs the same proxies for supply and oil price as Kilian (2009), but it differs in two ways. First, this study specifies all variables in levels instead using first differences. In contrast, Kilian (2009) specifies world crude oil production in percentage changes. Applying the variables at first differences eliminates the long-run relationship between oil production, demand, and oil price. Second, this study uses a different proxy for global demand. The use of freight rates as a common proxy for global demand has been discussed in the existing literature, mainly to assess variation in global economic growth (see Klovland (2004); Fan and Xu (2011)). Klovland (2004) finds an asymmetry relationship between the economic activity cycle and shipping freight rate. The close relationship between the business cycle, commodity prices, and freight rates is mainly due to the positive freight rate cycles. In addition to choosing the leading indicators for economic growth, Fan and Xu (2011) choose BDI over other proxies such as world oil inventory because BDI can capture the demand pressure between 2000 and 2009, a period when the world oil supply was relatively stable.

Both BDI and the real economic activity index proposed in Kilian (2009) are dry bulk ocean freight rate-based indices that are positively correlated, with the correlation value of 0.91. However, this section prefers BDI to Kilian's real economic activity index because BDI is available in level with original positive value. BDI is also available in a daily, weekly, and monthly frequency that is more flexible for analysing short- and long-term fluctuation. In contrast, Kilian (2009) constructs the index in growth rates then are then linearly detrended; this construction is more suitable for cyclical short-run relationships rather than for long-run relationships. Further, as it is a growth rate, Kilian's index consists of negative and positive values which are harder to interpret in the data transformation.

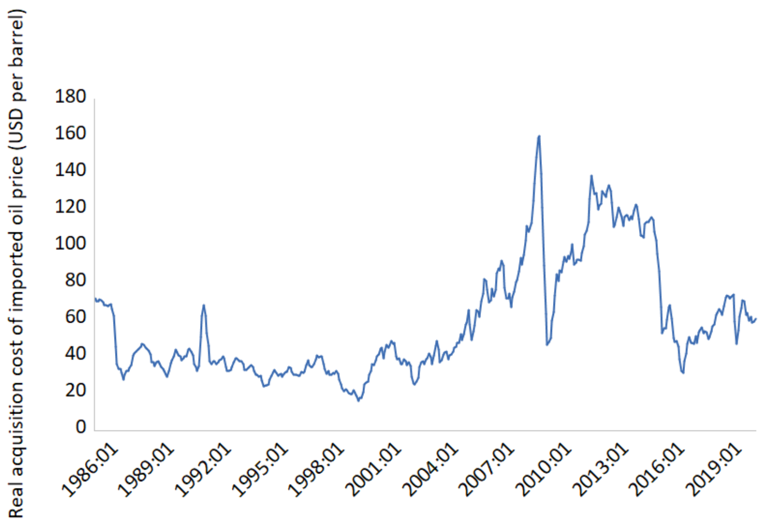


Figure 3.3: Crude oil price

Table 3.1: Descriptive statistics of crude oil market data

Variable	Mean	Median	Min	Max	Standard deviation	Skewness	Kurtosis
prod	69027.85	68745.75	51,324.41	84611.27	8318.348	-0.026	1.887
bdi	2680.714	2217.607	344.413	13,725.52	2025.344	2.577	11.209
op	57.842	45.534	15.634	159.686	31.068	1.072	3.184
*lprod**	11.135	11.138	10.846	11.346	0.122	-0.190	1.948
*lbdi**	7.691	7.704	5.842	9.527	0.619	0.225	3.515
*lop**	3.928	3.818	2.749	5.073	0.501	0.320	2.239

* *l* in variable denotes logged level.

3.3.2 Descriptive analysis of data

The time series plot of crude oil production is illustrated in Figure 3.1. World crude oil production exhibits an increasing trend throughout the observations from 1985 to 2019. Figure 3.2 shows that BDI is relatively stable, with few fluctuations. There is an extreme peak in October 2007 and May 2008 and a sudden drop in October 2008, which occurred in the Global Financial Crisis period. During the Global Financial Crisis, the global demand for commodities dropped, which can be expected to lower the shipping cost, as represented by the low index level. As illustrated in Figure 3.3, crude oil price fluctuates over time, mainly exhibiting peaks in July 2008 and extreme drops in

December 1998, December 2008, and February 2016. Fan and Xu (2011) argue that before June 2008, the main drivers of the oil price fluctuation are bubble accumulation, speculation, and episodic events due to attacks or wars. After June 2008, supply and demand play an essential role in oil price fluctuation. Ratti and Vespignani (2015) support the argument that strong global demand is the primary driver of high oil prices between 1997 and 2012. Baumeister and Kilian (2016*b*) conclude that the oil price decline after June 2014 is caused by a positive supply, weakening global demand, and an unpredictable component due to oil price expectation shocks.

Table 3.1 illustrates the descriptive statistics of the variables. None of the variables shows symmetrical distribution, as indicated by the non-zero values in their skewness. Taking a natural logarithm makes the data less skewed by pulling the high extreme values closer to the median, leading to nearly zero skewness value and reducing the heavy tails. The BDI and crude oil price are positively skewed; both variables have extremely high values. They have long right tails relative to the left tails, which means a large number of low indices and oil prices, and there are only a few values of high indices and prices. In contrast, world oil production is negatively skewed in that production volumes have a left tail that is slightly longer relative to the right tail, indicating that high oil production has a larger frequency than low oil production. The kurtosis value of oil production is light-tailed relative to the normal distribution, as indicated by its kurtosis value of less than three. The BDI and crude oil price are heavy-tailed relative to the normal distribution. The large kurtosis value indicates more outliers, extreme values of BDI, and real oil prices that are further away from the sample mean.

3.3.3 Unit root tests

This section applies the standard Augmented Dickey-Fuller (ADF) test and a test proposed by Enders and Lee (2012) to accommodate the possible breaks in the variables. The standard ADF test is described in Dickey and Fuller (1979). This section applies ADF by including a constant and linear time trend specification, as follows:

$$\Delta y_t = a_0 + \alpha\, y_{t-1} + a_2 t + \beta_1 \Delta y_{t-1} + ... + \beta_p \Delta y_{t-p} + \epsilon_t \tag{3.1}$$

where a_0 is the constant, t is the time trend, α and a_2 are the estimated parameters, p is the chosen lag length, β is the coefficient of the lagged independent variable, and ϵ_t is the white noise. The joint hypothesis of the unit root to be tested is $a_0 = \alpha = a_2 = 0$.

The existing studies commonly apply standard ADF for testing unit root. However, the drawback of the ADF test is that it tends to reject the null hypothesis of unit root when the series contains a structural break, which may lead to spurious regression. From a visual inspection, the BDI and real oil price indicate possible breaks. Thus, the unit root test with structural break is applied to overcome the inaccurate rejection of the null hypothesis. The most suitable global oil market model is then determined, as follows.

3.3.3.1 Enders and Lee (2012) test

Enders and Lee (2012) develop a unit root test that utilises the Fourier approximation to capture structural breaks in the series. The basic concept of this approach is that it modifies the ADF test with a deterministic term that can replicate the type of break. One of the benefits of this test is it can capture breaks without knowing *a priori* the form of the breaks. This deterministic term is in the form of trigonometric components, with a small number of frequency components that can capture the non-linear trend, gradual breaks, and a sharp break. The test's other benefit is that it requires no assumption about the number of breaks and the break dates. Hence, the test uses only a few parameters and it minimises power loss compared to the traditional unit root test with dummy variables (Becker et al., 2004; Enders and Lee, 2012).

Fourier approximation begins with the deterministic term $d(t)$ in a modification of the Dickey-Fuller (DF) unit root test in Dickey and Fuller (1979). Enders and Lee (2012) express the data generating process (DGP) and its subsequent regression based on a single frequency. In this study, multiple frequencies are applied to better represent an unknown deterministic term function. Then, Equation 3.3 describes Fourier expansion in deterministic term $d(t)$ with multiple frequencies,

$$y_t = d(t) + \rho y_{t-1} + \gamma t + \varepsilon_t \tag{3.2}$$

$$d(t) = \alpha_0 + \sum_{k=1}^{n} \alpha_k \, sin\Big(\frac{2\pi kt}{T}\Big) + \sum_{k=1}^{n} \beta_k \, cos\Big(\frac{2\pi kt}{T}\Big); \;\; n \leq \frac{T}{2} \tag{3.3}$$

where α_0 is the constant, α_k and β_k measure amplitude and displacement of the sinusoidal component of deterministic term, n is the number of cumulative frequency, and T is the total number of observations. Substituting Equation 3.3 into Equation 3.2 yields Equation 3.4.

$$y_t = \alpha_0 + \gamma t + \sum_{k=1}^{n} \alpha_k \, sin\Big(\frac{2\pi kt}{T}\Big) + \sum_{k=1}^{n} \beta_k \, cos\Big(\frac{2\pi kt}{T}\Big) + e_t \tag{3.4}$$

$$e_t = \rho \, e_{t-1} + \varepsilon_t \tag{3.5}$$

The null hypothesis is that the unit root is presence $\rho = 1$ with alternative hypothesis $\rho < 1$. The parameters of $\alpha_0, \gamma, \alpha_k$, and β_k do not depend whether the null hypothesis is true ($\rho = 1$) or not. Then, the distribution of the tests under null and alternative hypotheses are invariant to these four parameters. By imposing this null hypothesis restriction, the test statistic is obtained by estimating the model in first differences, as shown in Equation 3.6.

$$\Delta y_t = \delta_0 + \sum_{k=1}^{n} \delta_k \, \Delta sin\Big(\frac{2\pi kt}{T}\Big) + \sum_{k=1}^{n} \zeta_k \, \Delta cos\Big(\frac{2\pi kt}{T}\Big) + u_t \tag{3.6}$$

The detrended series $\hat{S}_t$ is constructed using estimated coefficients $\hat{\delta}_0, \hat{\delta}_k$, and $\hat{\zeta}_k$,

$$\hat{S}_t = y_t - \hat{\psi} - \hat{\delta}_0 t - \sum_{k=1}^{n} \hat{\delta}_k \, sin\Big(\frac{2\pi kt}{T}\Big) - \sum_{k=1}^{n} \hat{\zeta}_k \, cos\Big(\frac{2\pi kt}{T}\Big) \tag{3.7}$$

where

$$\hat{\psi} = y_1 - \hat{\delta}_0 - \sum_{k=1}^{n} \hat{\delta}_k \, sin\Big(\frac{2\pi k}{T}\Big) - \sum_{k=1}^{n} \hat{\zeta}_k \, cos\Big(\frac{2\pi k}{T}\Big). \tag{3.8}$$

As y_1 is the first observation of y_t, subtracting y_t with $\hat{\psi}$ results in $\hat{S}_1 = 0$. The regression of the unit root test with structural break to obtain the Lagrange Multiplier (LM) t-statistic is obtained through detrended series in Equation 3.7.

$$\Delta y_t = \phi \hat{S}_{t-1} + d_0 + \sum_{k=1}^{n} d_k \, \Delta sin\Big(\frac{2\pi kt}{T}\Big) + \sum_{k=1}^{n} z_k \, \Delta cos\Big(\frac{2\pi kt}{T}\Big) + \varepsilon_t \tag{3.9}$$

The nonstationary case is when $\phi = 0$ and the LM test statistic is given by

$$\tau_{LM} = t - \text{statistic for null hypothesis}\, \phi = 0. \tag{3.10}$$

3.3.3.2 Unit root tests result

Table 3.2 reports the results from the ADF and Enders and Lee (EL) unit root tests with multiple frequencies ($n = 3$). The results show that the variables in level have a mixed order of integration at the 5% significance level. Both tests confirm that world oil production in levels and log-levels are stationary. The time series visual inspection shows that world oil production exhibits an increasing trend and no possible breaks. Both tests reject the null hypothesis of unit root for BDI in levels, except that the ADF test fails to reject BDI in log-levels. Then, both tests fail to reject the null hypothesis of unit root for the crude oil price. The BDI and crude oil price indicate the possibility of the breaks, particularly during 2008–09 and 2014–16, as shown in Figures 3.2 and 3.3, respectively. Applying the EL test for the series with the possibility of the break is more powerful than the standard ADF unit root test. All variables are stationary in first differences, indicating that the maximum order of integration is order one.

Table 3.2: Unit root test for world crude oil production, BDI, and crude oil price

Variable	ADF				EL			
	Level	Log-level	first difference	Log-first difference	Level	Log-level	first difference	Log-first difference
prod	-5.418***	-5.365***	-6.014***	-6.161***	-5.686**	-5.647**	-25.714***	-25.620***
bdi	-3.490**	-2.907	-12.594***	-9.792***	-5.996***	-6.523***	-12.510***	-17.041***
op	-2.702*	-2.612*	-10.801***	-15.783***	-4.339	-3.508	-11.102***	-13.422***

*, **, and *** p-value is significant at 10%, 5%, and 1% significance level, respectively.

3.4 Methodology

This section illustrates the empirical framework for analysing the long-run equilibrium relationship between world oil production, BDI, and crude oil price. The ARDL approach is applied as it gives flexibility for dealing with stationary and nonstationary variables. First, this section presents the basic ARDL model in three single ARDL equations, with world oil production, BDI, and real oil price as the dependent variables. The aim is to examine the consistency of the long-run equilibrium presence and the sign of the response between variables. The ARDL in error-correction form is presented to give a more straightforward interpretation of world oil production, BDI, and oil price response to the global oil market variables. It distinguishes the short- and long-run equilibrium relationship. The interaction between variables might fluctuate in the short run, which causes mixed sign responses to the previous lags. If there is cointegration, the explanatory variables are, in the long run, forcing the dependent variable into equilibrium. Thus, the long-run equilibrium relationship is a helpful way of interpreting the relationship among the variables.

Second, this section presents the asymmetric effects of the global oil market. The supply, aggregate demand, and crude oil price fluctuations decompose into partial positive and negative changes, from which it may be understood whether positive or negative shocks play an important role in supply, demand, and crude oil price. Then, the asymmetric result in the error-correction representation is also presented to distinguish long-run and short-run asymmetric effects.

3.4.1 The ARDL and ARDL in Error-Correction form (ARDL-EC)

The ARDL equation follows Equation 2.41, while the error correction form follows Equation 2.42, each with two explanatory variables ($k = 2$).

$$y_t = a_0 + a_1 t + \sum_{i=1}^{p} \phi_i \, y_{t-i} + \sum_{j=1}^{k} \sum_{i=0}^{q} \beta'_{j,i} \, x_{j,t-i} + \epsilon_t \tag{2.41}$$

The y_t is the dependent variable, $x_1, ..., x_k$ are k explanatory variables, with p and q being the lags of the dependent and explanatory variables, respectively. The optimal lags p and q are obtained by minimising the value of information criteria AIC, as described in Equation 2.26. The constant term is a_0, the error term is ϵ_t, and the respective coefficients a_1, ψ_i, β_{j,l_j} are the coefficients of the linear trend t, lags of y_t, and lags of the k explanatory variables $x_{j,t}$.

The empirical study in this chapter runs three single ARDL equations for world oil production, BDI, and crude oil price. In the first ARDL equation, y_t is world oil production (the dependent variable), and x_1, x_2 are the BDI and crude oil price (the regressors). The second ARDL equation applies BDI as y_t, with the other variables of world oil production and crude oil price being x_1, x_2. The third equation applies crude oil price as the dependent variable y_t, and world oil production and BDI are the regressors x_1, x_2. The world oil production, BDI, and oil price equations follow case five of the deterministic terms, namely unrestricted intercept and unrestricted time trend. Based on graphical interpretation, world oil production exhibits an increasing time trend t. Empirically, the time trend presence for each series is tested using linear regression, and it shows a statistically significant trend presence for world oil production. The ADF test with trend specification confirms that world oil production and crude oil price equations exhibit a significant deterministic trend. World oil production and real oil price also have a positive and significant time-trend, but have low magnitude in the short-run ARDL model.

The regression expects that crude oil production negatively relates to crude oil price and positively relates to BDI. BDI relates positively to crude oil price and production. The crude oil price relates negatively to supply, which means a drop in crude oil production is associated with the oil price. Crude oil price relates positively to global demand.

$$\Delta y_t = a_0 + a_1 t - \alpha(y_{t-1} - \theta x_{t-1}) + \sum_{i=1}^{p-1} \psi_{yi}\, \Delta y_{t-i} + \omega^{'} \Delta x_t + \sum_{i=1}^{q-1} \beta^{'}_{xi}\, \Delta x_{t-i} + \epsilon_t \tag{2.42}$$

The speed-of-adjustment α represents how much the dependent variable needs to adjust from the previous period to revert to equilibrium in the current period, and it is given by $\alpha = 1 - \sum_{j=1}^{p} \phi_i$ where ϕ refers to the coefficient of the lagged dependent variable in Equation 2.41. As the test distribution is not standard, the critical values of the speed-of-adjustment follow Kripfganz et al. (2018). The long-run coefficients θ is given by $\theta = \frac{\sum_{j=0}^{q} \beta_j}{\alpha}$ and represents the effects of the independent variables on the dependent variable in equilibrium. The short-run coefficients, ψ, $\omega^{'}$, $\beta^{'}$ indicate short-term fluctuations of the explanatory variables on the dependent variables.

The long-run equilibrium or cointegration evidence requires both F-statistics and t-statistics to be rejected. The bounds testing calculates the F-statistics of the joint null hypothesis that the speed of adjustment and the coefficients of the lagged explanatory variables are zero $H_0^F : (\alpha = 0)\big(\sum_{j=0}^{q} \beta_j = 0\big)$. The F-statistics are then compared to the critical values, which follows Kripfganz et al. (2018) for the finite-sample and asymptotic critical values that provide more efficient estimates, regardless of the lag length or the number of short-run coefficients. If the F-statistics is rejected, then the value of t-statistics needs to be checked. The null hypothesis is that the coefficient of the speed of adjustment is zero $H_0^t : \alpha = 0$.

Equations 3.11 and 3.12 describe the ARDL and ARDL in error correction form models with a structural break. A dummy variable (b) represents the break in crude oil price, global demand, and supply incorporated in the ARDL model. It takes a value of 1 from the estimated break date onwards and a value of 0 from the beginning of observation until before the estimated break date. The estimated breakpoint for each equation is obtained based on the supremum Wald test by Andrews (1993). If the estimated break date for three equations results in different dates, the date on which the strongest jump occurred in the data series will be used as a dummy variable. The breaks in intercept and slopes are considered, with the coefficient estimates γ_0 for a break in intercept and γ_j for the break in slopes. The interaction of the break in slopes is denoted by bx_t, which is the multiplication of the break date dummy b and the regressors x_t.

$$y_t = a_0 + a_1 t + \sum_{i=1}^{p} \phi_i \, y_{t-i} + \sum_{j=1}^{k} \sum_{i=0}^{q} \beta'_{j,i} \, x_{j,t-i} + \gamma_0 b_t + \sum_{j=1}^{k} \gamma_j \, bx_t + \epsilon_t \quad (3.11)$$

$$\begin{aligned} \Delta y_t = a_0 + a_1 t - \alpha(y_{t-1} - \theta x_{t-1}) + \sum_{i=1}^{p-1} \psi_{yi} \, \Delta y_{t-i} + \omega' \Delta x_t + \\ \sum_{i=1}^{q-1} \beta'_{xi} \, \Delta x_{t-i} + \gamma_0 b_t + \sum_{j=1}^{k} \gamma_j \, bx_t + \epsilon_t \end{aligned} \quad (3.12)$$

The Supremum Wald test statistics W_T is given by Equation 3.13.

$$Supremum \; W_T = \sup_{\beta_1 \leq \pi \leq \beta_2} W_T(\pi), \quad (3.13)$$

where π is the break-point and lies within the range of β_1 and β_2. The null hypothesis is no break or the stability of the parameter,

$$H_0 : \beta_t = \beta_0 \quad (3.14)$$

and the alternative hypothesis is

$$H_1 : \beta_t = \begin{cases} \beta_1(\pi)\ for\ t = 1, .., T_\pi \\ \beta_2(\pi)\ for\ t = T_{\pi+1}, ..., T. \end{cases} \tag{3.15}$$

The limiting distributions of the test statistics are

$$Supremum\ W_T \rightarrow_d \sup_{\lambda \in [\varepsilon_1, \varepsilon_2]} W(\lambda) \tag{3.16}$$

3.4.2 Nonlinear Autoregressive Distributed Lags (NARDL)

Shin et al. (2013) proposes NARDL(p, q) by extending the basic ARDL model as proposed by Pesaran and Shin (1997) and Pesaran et al. (2001) to incorporate a feature that distinguishes the asymmetric effects of the explanatory variables. The asymmetric effects disentangle the explanatory variables into lags of the dependent variable and lags of the positive and negative growth of the regressors. The NARDL model identifies the positive or negative changes in world oil production, BDI, and crude oil price in the following equation:

$$y_t = a_0 + a_1 t + \sum_{i=1}^{p} \phi_i\, y_{t-i} + \sum_{j=1}^{k} \sum_{i=0}^{q} \left(\beta_{j,i}^{+'} x_{t-i}^{+} + \beta_{j,i}^{-'} x_{t-i}^{-} \right) + \varepsilon_t \tag{3.17}$$

where x_t is a k x 1 vector of regressors: $x_t = x_0 + x_t^+ + x_t^-$. The partial positive is defined as the lagged value of the partial positive added by the positive changes, and the partial negative is the lagged value of partial negative added by the negative changes (or subtracted by the decrease from the previous to current period). NARDL in error correction (NARDL-EC) representation is defined by Equation 3.18.

$$\Delta y_t = a_0 + a_1 t - \alpha\big(y_{t-1} - \sum_{j=1}^{k} \theta_j^{+'} x_{t-1}^{+} - \sum_{j=1}^{k} \theta_j^{-'} x_{t-1}^{-}\big) + \sum_{i=1}^{p-1} \psi_{y,i} \Delta y_{t-i}$$
$$+ \sum_{j=1}^{k} \big(\omega_j^{+'} \Delta x_t^{+} + \omega_j^{-'} \Delta x_t^{-}\big) + \sum_{j=1}^{k} \sum_{i=1}^{q-1} \big(\psi_{xj,i}^{+'} \Delta x_{t-i}^{+} + \psi_{xj,i}^{-'} \Delta x_{t-i}^{-}\big) + \varepsilon_t \quad (3.18)$$

$$\Delta y_t = a_0 + a_1 t - \alpha \xi_{t-1} + \sum_{i=1}^{p-1} \psi_{y,i} \Delta y_{t-i} + \sum_{j=1}^{k} \big(\omega_j^{+'} \Delta x_t^{+} + \omega_j^{-'} \Delta x_t^{-}\big)$$
$$+ \sum_{j=1}^{k} \sum_{i=1}^{q-1} \big(\psi_{xj,i}^{+'} \Delta x_{t-i}^{+} + \psi_{xj,i}^{-'} \Delta x_{t-i}^{-}\big) + \varepsilon_t \quad (3.19)$$

where $\alpha = 1 - \sum_{i=1}^{p} \phi_i$, $\psi_{yi} = -\sum_{i=l+1}^{p} \phi_i$ for $l = 1, ..., p-1$, $\beta_{j,i}^{+'} = \sum_{i=0}^{q} \beta_{j,i}^{+'}$, $\beta_{j,i}^{-'} = \sum_{i=0}^{q} \beta_{j,i}^{-'}$, $\omega_j^{+'} = \beta_{j,0}^{+'}$, $\omega_j^{-'} = \beta_{j,0}^{-'}$, $\psi_{xj,i}^{+'} = -\sum_{i=l+1}^{q} \beta_{j,i}^{+'}$, for $l = 1, ..., q-1$, $\psi_{xj,i}^{-'} = -\sum_{i=l+1}^{q} \beta_{j,i}^{-'}$, for $j = 1, ..., q-1$, and for $k = 2$, the non-linear error correction term is $\xi_{t-1} = y_{t-1} - \theta_1^{+'} x_{t-1}^{+} - \theta_2^{+'} x_{t-1}^{+} - \theta_1^{-'} x_{t-1}^{-} - \theta_2^{-'} x_{t-1}^{-}$, where $\theta_j^{+'} = \beta_{j,1}^{+'}/\alpha$ and $\theta_j^{-'} = \beta_{j,1}^{-'}/\alpha$ are the asymmetric long-run parameters.

The NARDL model with a structural break is expressed in Equations 3.20 and 3.21, for NARDL and NARDL-EC with the break, respectively.

$$y_t = a_0 + a_1 t + \sum_{i=1}^{p} \phi_i \, y_{t-i} + \sum_{j=1}^{k} \sum_{i=0}^{q} \big(\beta_{j,i}^{+'} x_{t-i}^{+} + \beta_{j,i}^{-'} x_{t-i}^{-}\big) + \gamma_0 b_t$$
$$+ \sum_{j=1}^{k} \big(\gamma_j^{+'} \, b x_t^{+} + \gamma_j^{-'} \, b x_t^{-}\big) + \varepsilon_t, \quad (3.20)$$

where $\gamma_j^{+'}$ and $\gamma_j^{-'}$ denote the coefficient estimates of the positive and negative shocks of the independent variables, respectively.

$$\Delta y_t = a_0 + a_1 t - \alpha\big(y_{t-1} - \sum_{j=1}^{k} \theta_j^{+'} x_{t-1}^{+} - \sum_{j=1}^{k} \theta_j^{-'} x_{t-1}^{-}\big) + \sum_{i=1}^{p-1} \psi_{y,i} \Delta y_{t-i}$$
$$+ \sum_{j=1}^{k} \left(\omega_j^{+'} \Delta x_t^{+} + \omega_j^{-'} \Delta x_t^{-}\right) + \sum_{j=1}^{k} \sum_{i=1}^{q-1} \left(\psi_{xj,i}^{+'} \Delta x_{t-i}^{+} + \psi_{xj,i}^{-'} \Delta x_{t-i}^{-}\right) + \varepsilon_t \tag{3.21}$$

$$\Delta y_t = a_0 + a_1 t - \alpha \xi_{t-1} + \sum_{i=1}^{p-1} \psi_{y,i} \Delta y_{t-i} + \sum_{j=1}^{k} \left(\omega_j^{+'} \Delta x_t^{+} + \omega_j^{-'} \Delta x_t^{-}\right)$$
$$+ \sum_{j=1}^{k} \sum_{i=1}^{q-1} \left(\psi_{xj,i}^{+'} \Delta x_{t-i}^{+} + \psi_{xj,i}^{-'} \Delta x_{t-i}^{-}\right) + \gamma_0 b_t + \sum_{j=1}^{k} \left(\gamma_j^{+'} b x_t^{+} + \gamma_j^{-'} b x_t^{-}\right) + \varepsilon_t \tag{3.22}$$

The interaction between break and positive shock of the independent variable is denoted with bx_t^{+}, whilst bx_t^{-} represents the break interaction with the negative shock of the independent variable.

3.5 Empirical result: the relationship between world oil production, demand, and crude oil price

This section reports the results of the estimations of the ARDL models described in Section 3.4. The basic ARDL result is presented in Subsection 3.5.1 (referring to Equation 2.41), with the error-correction representation in Subsection 3.5.1.1 (referring to Equation 2.42). The results of the asymmetric ARDL extension to the basic model are reported in Subsection 3.5.2, with the error-correction representation in Subsection 3.5.2.1.

3.5.1 The basic ARDL results

The optimal lag lengths of the variables for each of the three ARDL equations are chosen based on AIC. The first equation, for world oil production, has an optimal lag length of ARDL(1,0,3), with lag one for world production as the dependent variable (lprod), and lags zero and three for BDI (lbdi) and crude oil price (lop), respectively, as the independent variables. The optimal lag length for the second equation is ARDL(2,0,2), with BDI as the dependent variable with lag two. World oil production and crude oil price are the independent variables with lags zero and two, respectively. The third equation is the crude oil price equation with optimal lag length ARDL(3,1,1). Crude oil price is the dependent variable with lag three, while world oil production and BDI, as independent variables, both have lag one. Although the optimal lag length according to AIC differs across three equations, this chapter aligns the lag structure across the three equations by setting the lag length to be the longest suggested by AIC for each equation so that the three equations are consistent. For the three equations, the longest lags are lag one for world oil production, lag two for BDI, and lag three for the crude oil price. Thus, the final world oil production equation applies ARDL(1,2,3), while the BDI equation applies ARDL(2,1,3), and the crude oil price equation applies ARDL(3,1,2).

Table 3.3: ARDL estimation for world oil production, BDI, and crude oil price

	lprod_t	lbdi_t	lop_t
Coefficient estimates ($\phi_i, \beta'_{j,i}$)			
lprod_t		0.690 (0.856)	-1.032*** (0.336)
lprod_{t-1}	0.842*** (0.026)	-0.164 (0.853)	0.596* (0.337)
lbdi_t	0.002 (0.003)		0.056*** (0.020)
lbdi_{t-1}	-0.003 (0.004)	1.043*** (0.050)	-0.010 (0.028)
lbdi_{t-2}	0.004 (0.003)	-0.118** (0.050)	-0.025 (0.020)
lop_t	-0.022*** (0.007)	0.356*** (0.124)	
lop_{t-1}	0.022* (0.012)	-0.094 (0.216)	1.408*** (0.050)
lop_{t-2}	0.011 (0.012)	-0.197 (0.216)	-0.554*** (0.081)
lop_{t-3}	-0.013* (0.007)	-0.054 (0.123)	0.106** (0.048)
(a_1) t	1.7e-04*** (2.8e-05)	-7e-04 (5e-04)	5.7e-04*** (2e-04)
(a_0) Constant	1.666*** (0.277)	-4.979 (4.996)	4.551** (1.971)
Observations	418	418	418
F-statistic	6315.86	517.68	2295.76
Prob (F-statistic)	0.0000	0.0000	0.0000
Adjusted R^2	0.9927	0.9179	0.9803
Root MSE	0.0103	0.1779	0.0706

Notes:
Standard errors are in parentheses * $p < .10$, ** $p < .05$, *** $p < .01$

Table 3.3 summarises the results from estimating the three ARDL equations for world oil production, BDI, and the crude oil price. The first column of Table 3.3 presents the model estimated with the log of oil production as a dependent variable. The coefficient of crude oil price in the current period is significantly negative and the coefficients of the lags of oil production and time trend are significantly positive. The coefficients of oil price and the lag of BDI are significantly positive in the second column, which is the model with the log of BDI as the dependent variable. In the third column, where oil price is the dependent variable, the coefficient of oil production is significantly negative and the coefficients of BDI, time trend, and lag of oil price are significantly positive.

Economic theory suggests that quantity demanded decreases along the demand curve when the price rises. As the price rises, quantity supplied also increases along its own curve. When the demand curve expands, it shifts to the right and causes a price increase, whereas when the supply curve expands to the right, the price decreases. A crude oil price increase is associated with a decrease in world oil production and an increase in global demand. In the current period, a 1% rise in crude oil price significantly decreases oil production by 0.02% and increases the global demand by 0.36%. The increase in global demand leads to higher oil production and prices. A 1% rise in global demand increases world oil production and real oil price by 0.002% and 0.06%, respectively. The world oil production increase causes a decline in oil price and increases global demand. A 1% rise in crude oil production causes the real oil price to decline by 1.03%, increasing global demand by 0.69%.

Table 3.4: Structural break test result for world crude oil production, BDI, and crude oil price

Equation	Supremum Wald statistic	Estimated break date
lprod	45.3497***	April 1990
lbdi	71.9102***	January 2012
lop	46.1463***	January 2009

Notes:
Standard error in parentheses; * $p < .10$, ** $p < .05$, *** $p < .01$
The critical value of Wald test is based on Hansen (1997).

Table 3.4 shows the structural break test applying supremum Wald for the three equations. The first column refers to the dependent variable in each ARDL equation, i.e. lprod indicates the log of oil production as the dependent variable. The supremum Wald statistics for all equations reject the null hypothesis of no structural break, suggesting the coefficients in the regression vary over time. The supremum Wald tests estimate various break dates for each equation; in April 1990 for the production equation, January 2012 for the BDI equation, and January 2009 for the oil price equation. Based on visual interpretation, the time series plots of BDI and real oil prices, shown in Figures 3.2 and 3.3, depict extreme peaks and drops in mid-2008, while the crude oil production time series (refer to Figure 3.1) shows an increasing trend without extreme shifts. The estimated break date in January 2009 is the closest one following the extreme episode in mid-2008. The following model applies a structural break taking into account a break date in January 2009 following big shocks due to Global Financial Crisis. The dummy variable takes a value of 0 from January 1985 to December 2008 and 1 from January 2009 to December 2019.

Table 3.5: ARDL with a structural break model for world oil production, BDI, and crude oil price

	lprod_t	lbdi_t	lop_t
Coefficient estimates $(\phi_i, \beta'_{j,i}, \gamma_0, \gamma_j)$			
lprod_t		0.807 (0.849)	-0.900*** (0.327)
lprod_{t-1}	0.833*** (0.028)	-0.448 (0.849)	0.495 (0.331)
lbdi_t	0.003 (0.003)		0.080*** (0.020)
lbdi_{t-1}	-0.003 (0.004)	1.019*** (0.050)	-0.013 (0.028)
lbdi_{t-2}	0.003 (0.003)	-0.139*** (0.050)	-0.013 (0.020)
lop_t	-0.022*** (0.008)	0.356*** (0.125)	
lop_{t-1}	0.022* (0.013)	-0.060 (0.214)	1.362*** (0.049)
lop_{t-2}	0.011 (0.013)	-0.184 (0.213)	-0.509*** (0.079)
lop_{t-3}	-0.013* (0.007)	-0.063 (0.122)	0.076 (0.047)
b_t	0.014 (0.019)	16.964*** (5.855)	9.095*** (2.141)
b*lprod_t		-1.472*** (0.508)	-0.759*** (0.185)
b*lbdi_t	-0.002 (0.003)		-0.068*** (0.017)
b*lop_t	3.7e-04 (0.003)	-0.115* (0.061)	
(a_1) t	1.7e-04*** (3e-05)	-2.77e-04 (5.5e-04)	4.96e-04** (2e-04)
(a_0) Constant	1.757*** (0.293)	-3.099 (5.126)	4.100** (2.014)
Observations	417	417	417
F-statistic	4715.48	398.21	1834.32
Prob (F-statistic)	0.0000	0.0000	0.0000
Adjusted R^2	0.9927	0.9197	0.9814
Root MSE	0.0103	0.1759	0.0684

Notes:
Standard errors are in parentheses * $p < .10$, ** $p < .05$, *** $p < .01$

The ARDL model with a break date in January 2009 is presented in Table 3.5. The break in intercept indicated by b_t is statistically significant in the equations, with BDI and oil price as the dependent variable. Furthermore, the breaks in slope also show the statistical significance in BDI and oil price equations. The presence of the break in the slopes has negative effects on oil prices and global demand. The break interactions with global demand and oil production strongly affect crude oil prices, indicated by the statistical significance of b*lbdi_t and b*lprod_t in Table 3.5. The break in oil production and price also significantly affect global demand, denoted by the significance of b*lprod_t and b*lop_t in the equation with BDI as the dependent variable. The break in oil price and demand has little non-significant effects on oil production. The little effect of the breaks can be explained by the long lags between oil discovery and production start date. The following error correction form and asymmetric ARDL retain the break in the model to compare the significance and consistency with the basic model without a structural break.

3.5.1.1 ARDL in error correction interpretation

Table 3.6 reports the bounds test result, adjustment factor, long-run equilibrium, and short-run dynamics coefficients. The Pesaran et al. (2001) bounds test result indicates that a long-run equilibrium relationship is present among world oil production, BDI, and crude oil price for all three equations. The top section of Table 3.6 reports that the F-statistic and t-statistic of the adjustment factors are larger than the bounds test critical values at a 5% significance level for three equations. The null hypothesis that the level relationship does not exist is rejected; hence there is enough evidence to support the long-run relationship at the level of crude oil production, BDI, and real oil price.

The long-run equilibrium relationship denoted by coefficient θ in Table 3.6 represents a contemporaneous relationship among world oil production,

Table 3.6: ARDL estimation in error correction form for world oil production, BDI, and crude oil price

	$D.lprod_t$	$D.lbdi_t$	$D.lop_t$
(i) *Case 5*			
Bound Test H_0: no level relationship			
F-stat‡	12.129***	6.434**	6.181**
t-stat‡	-6.014***	-4.083**	-4.023**
(ii) *Adjustment factor (-α)*‡			
$lprod_{t-1}$	-0.158*** (0.026)		
$lbdi_{t-1}$		-0.075*** (0.018)	
lop_{t-1}			-0.039*** (0.010)
(iii) *Long-run* (θ)			
$lprod_{t-1}$		7.047 (5.785))	-11.144** (4.774)
$lbdi_{t-1}$	0.020*** (0.006)		0.544*** (0.179)
lop_{t-1}	-0.015* (0.009)	0.152 (0.320)	
(iv) *short-run* ($\psi_{yi}, \omega', \psi'_{xi}$)			
$D.lprod_t$		0.690 (0.856)	-1.032*** (0.336)
$D.lbdi_t$	0.002 (0.003)		0.056*** (0.019)
$D.lbdi_{t-1}$	-0.004 (0.003)	0.118** (0.050)	0.025 (0.020)
$D.lop_t$	-0.022*** (0.007)	0.356*** (0.124)	
$D.lop_{t-1}$	0.002 (0.008)	0.251* (0.134)	0.447*** (0.048)
$D.lop_{t-2}$	0.013* (0.007)	0.054 (0.123)	-0.106** (0.048)
(a_1) t	1.7e-04*** (2.8e-05)	-7e-04 (5e-04)	5.7e-04*** (2e-04)
(a_0) Constant	1.666*** (0.277)	-4.979 (4.996)	4.551** (1.971)
Observations	418	418	418

Notes:
Standard errors are in parentheses * $p < .10$, ** $p < .05$, *** $p < .01$
‡The approximate p-values applied for speed-of-adjustment coefficient is based on Kripfganz et al. (2018)

BDI, and real oil price, how three variables are related in the long run. The consistency of sign and magnitude is checked across the three equations by converting two equations so that the three equations express the same variable (oil price) as a function of the other two variables (demand and supply). Equations 3.23-3.25 are the error correction forms derived from the cointegrating vectors in Table 3.6, Panel (iii). The cointegrating vector of the supply and demand equations is expressed in Equations 3.23 and 3.24, respectively. As an illustration, the second column of Panel (iii) gives production as the dependent variable, with BDI and oil price as independent variables in the long-run relationship. To obtain Equation 3.23, the equation in Panel (iii) is rearranged such that the oil price is on the left-hand side of the equation, with production and BDI on the right-hand side. That would be applied to BDI as the dependent variable in the third column of Panel (iii). Oil price is on the left-hand side while production and BDI are on the right-hand side to obtain Equation 3.24.Equation 3.25 is taken from the fourth column of Panel (iii) without rearrangement, as the oil price is already on the left-hand side. Hence, the magnitude of the three variables can be compared. For all three equations, the coefficient is positive and greater for demand; in contrast, it is negative and smaller for supply.

$$lprod_{t-1} = 0.020\, lbdi_{t-1} - 0.015\, lop_{t-1}$$

$$\Longleftrightarrow\ lop_{t-1} = -67.219\, lprod_{t-1} + 1.319\, lbdi_{t-1} \tag{3.23}$$

$$lbdi_{t-1} = 7.047\, lprod_{t-1} + 0.152\, lbdi_{t-1}$$

$$\Longleftrightarrow\ lop_{t-1} = -46.368\, lprod_{t-1} + 6.580\, lbdi_{t-1} \tag{3.24}$$

$$lop_{t-1} = -11.144\, lprod_{t-1} + 0.544\, lbdi_{t-1} \tag{3.25}$$

The coefficients in Equations 3.23-3.25 should be interpreted carefully, and the interpretation is helpful for the sign and magnitude consistency checking across three equations to understand how they relate each other in the long-run relationship. The cointegrating vector from the supply equation represents how supply relates to demand and oil price as presented in the first line of Equation 3.23 interpreted as a 1% increase in oil prices relate negatively with the production by 0.015% in the long run. Supply appears to exhibit a small response in percentage, but sample variance (0.015) relative to the sample mean for supply (11.144) is also small. Thus, a 1% response can be bigger in terms of the sample. Comparing the coefficient of production in Equations 3.23 and 3.25 (-67.219 and -11.144, respectively), Equation 3.23 requires price to adjust by a magnitude of six times larger to the supply shock.

The consistency check based on magnitudes shows that demand shock is associated with a greater price change compared to supply shock. Comparing the coefficient of BDI in Equations 3.24 and 3.25 (6.580 and 0.544, respectively), Equation 3.24 indicates that oil price requires to adjust by a magnitude of twelve times larger to the demand shock.

The adjustment factor ($-\alpha$) shows how the dependent variable changes when the three variables deviate from their long-run equilibrium. The adjustment factors for the three equations are all negative and statistically significant, which means that the estimated error correction forms represent stable relationships. The dependent variable of the prior period is too high relative to the long-run equilibrium and so it is necessary to decrease its value in the current period to revert to the equilibrium. When the world oil production is 1% higher than the level of the long-run equilibrium, it decreases by 0.16% in the subsequent period. In addition, if the BDI is 1% higher than its level at the long-run equilibrium, it decreases by 0.08% in the subsequent period. Finally, when the oil price is 1% higher than its level at the long-run equilibrium, it declines by 0.04% in the subsequent period. However, these adjustment factors are not directly comparable because the supply's 1% deviation from the long-run equilibrium is not identical to a 1% deviation in demand or oil price.

Equations 3.26-3.28 describe the adjustment factor across the three equations to enable direct comparison when considering a 1% deviation in the oil price, where sr refers to the short-run dynamics. The comparison of adjustment factors is interpreted carefully. The long-run equilibrium relationship is defined in terms of the dependent variable; for instance, Equation 3.26 tells the adjustment factor for the supply equation, which describes how much supply changes when log supply is one unit higher than the level at the long-run equilibrium.

Similarly, the adjustment factor for the price equation in Equation 3.28 shows how much price changes when the log oil price is one unit higher than the level at the long-run equilibrium. As an illustration, the adjustment factor in the first row of Equation 3.26 is derived from Panel (ii) of Table 3.6, with the long run coefficients from Panel (iii). The adjustment factor for price in the third row of Equation 3.28 is taken directly from Panel (ii) of Table 3.6 without any rearrangement. The Equation 3.26 is rearranged so that a unit of log price in the supply equation (Equation 3.26) can be compared with a unit of log price in the price equation (Equation 3.28). The three equations show that crude oil price adjusts more (the coefficient of -0.039 in Equation 3.28) when the three variables deviate from the long-run equilibrium. Supply (the coefficient of -0.002 in Equation 3.26) is the slowest in changing (0.06 times less than price) when the three variables deviate from long-run equilibrium. Demand (the coefficient of -0.011 in Equation 3.27) is faster to adjust than supply but is still 0.29 times slower than the oil price.

$$D.lprod_t = -0.158\,(lprod_{t-1} - 0.022\,lbdi_{t-1} + 0.014\,lop_{t-1}) + sr$$

$$\iff D.lprod_t = -0.158\,lprod_{t-1} + 0.004\,lbdi_{t-1} - 0.002\,lop_{t-1} + sr$$

$$\iff D.lprod_t = -0.002\,(lop_{t-1} + 79\,lprod_{t-1} - 2\,lbdi_{t-1}) + sr \tag{3.26}$$

$$D.lbdi_t = -0.075\,(lbdi_{t-1} - 7.047\,lprod_{t-1} - 0.152\,lop_{t-1}) + sr$$

$$\iff D.lbdi_t = -0.075\,lbdi_{t-1} + 0.529\,lprod_{t-1} + 0.011\,lop_{t-1}) + sr$$

$$\iff D.lbdi_t = -0.011\,(lop_{t-1} + 48.091\,lprod_{t-1} - 6.818\,lbdi_{t-1}) + sr \tag{3.27}$$

$$D.lop_t = -0.039\,(lop_{t-1} + 11.144\,lprod_{t-1} - 0.544\,lbdi_{t-1}) + sr \tag{3.28}$$

The short-run dynamic relationship indicates that crude oil price significantly affects world oil production and BDI. A 1% rise in real oil price increases the BDI by 0.36% and decreases the oil production by 0.02%. World oil production and BDI also exert statistically significant effects on the short-run crude oil price. A 1% increase in oil production decreases the real oil price by 1.03%, and a 1% BDI rise increases the crude oil price by 0.06%. The positive linear time trend is statistically significant for world oil production and crude oil price.

Table 3.7 presents ARDL model with a structural break in error correction form. The model with a break in January 2009 shows the presence of the long-run equilibrium relationship as indicated by the rejection of the Bounds test for all three equations in Panel (i) of Table 3.7. The results confirm consistency with an earlier model without a break in terms of the sign of the coefficients. The adjustment factor ($-\alpha$) in Panel (ii) is negative and statistically significant for all equations, which indicates the stable relationship for the dependent variable to respond when the three variables deviate from long-run equilibrium. The long-run coefficients (θ) also have consistent signs with the model without a break.

The sign of the long-run relationship between oil price, production, and demand varies as the ARDL approach in this thesis does not rely on the sign-identifying assumptions as applied by Kilian (2009); Kilian and Murphy (2014). Baumeister and Hamilton (2019) suggests relaxing the assumption on the sign restriction as it affects the magnitude of price elasticity on demand and supply. In the ARDL model, the long-run relationship does not represent causality between variables. However, it models the equilibrium relationships of demand, supply, and price after testing for the presence of long-run equilibrium and how the three variables move in response to the deviation from the long-run equilibrium. One of the possible reasons for the negative relationship between oil prices and production in the short run is that the rise of crude oil prices affects transportation costs, which may slow down oil production activity at the current period. The second reason is that it takes some time to adjust production volume. A high oil price will give the incentive to find more oil

Table 3.7: ARDL estimation with a structural break in error correction form for world oil production, BDI, and crude oil price

	D.lprod$_t$	D.lbdi$_t$	D.lop$_t$
(i) *Case 5*			
Bound Test H_0: no level relationship			
F-stat‡	12.133***	9.499***	14.155***
t-stat‡	-6.013***	-5.122***	-6.225***
(ii) *Adjustment factor (-α)*‡			
lprod$_{t-1}$	-0.168*** (0.028)		
lbdi$_{t-1}$		-0.119*** (0.023)	
lop$_{t-1}$			-0.072*** (0.012)
(iii) *Long-run* (θ)			
lprod$_{t-1}$		3.012** (3.993))	-5.657** (2.702)
lbdi$_{t-1}$	0.022** (0.010)		0.751*** (0.135)
lop$_{t-1}$	-0.014 (0.012)	0.412* (0.243)	
(iv) *short-run* ($\psi_{yi}, \omega', \psi'_{xi}, , \gamma_0, \gamma_j$)			
D.lprod$_t$		0.807 (0.849)	-0.900*** (0.327)
D.lbdi$_t$	0.003 (0.003)		0.080*** (0.020)
D.lbdi$_{t-1}$	-0.003 (0.003)	0.139*** (0.050)	0.013 (0.020)
D.lop$_t$	-0.022*** (0.007)	0.356*** (0.125)	
D.lop$_{t-1}$	0.002 (0.008)	0.247* (0.133)	0.434*** (0.047)
D.lop$_{t-2}$	0.013* (0.007)	0.063 (0.122)	-0.076 (0.047)
b$_t$	0.014 (0.019)	16.964*** (5.855)	9.095*** (2.141)
b*lprod$_t$		-1.472*** (0.508)	-0.759*** (0.185)
b*lbdi$_t$	-0.002 (0.003)		-0.068*** (0.017)
b*lop$_t$	3.7e-04 (0.003)	-0.115* (0.061)	
(a_1) t	1.7e-04*** (3e-05)	-2.77e-04 (5.5e-04)	4.96e-04*** (2e-04)
(a_0) Constant	1.757*** (0.293)	-3.099 (5.126)	4.100** (2.014)
Observations	417	417	417

Notes:
Standard errors are in parentheses * $p < .10$, ** $p < .05$, *** $p < .01$
‡The approximate p-values applied for speed-of-adjustment coefficient is based on Kripfganz et al. (2018)

through exploration and it takes a few years to start producing from new fields. The production volume adjustment depends on the existing geographical condition (i.e. existing fields' capacity) and could also be affected by worldwide geopolitical issues considering world data sets are applied in this study.

3.5.2 The asymmetric ARDL (NARDL) results

Asymmetry represents the difference between a positive and negative change in world oil production, BDI, and crude oil price. The variables in the basic ARDL model are decomposed into partial positive and negative changes in the NARDL model. The optimal lag lengths for each equation follow the basic ARDL model as described in Subsection 3.5.1. In the first equation, world oil production has the lag length of NARDL(1,2,2,3,3): lag one for world oil production as the dependent variable, and lags two and three for the positive and negative changes of BDI and crude oil prices respectively, as the explanatory variables. Copying across those same lag lengths for the variables, the BDI equation follows NARDL(2,1,1,3,3) with lag two for the BDI as the dependent variable and lags one and three for the regressors, world oil production and crude oil price. The crude oil price equation follows NARDL(3,1,1,2,2) with lag three for oil price and lags one and two for the positive and negative changes in oil production and BDI, respectively.

Table 3.8 presents the results from estimating NARDL for the three equations. In the first column of the table, the estimated NARDL for world oil crude production shows that the partial positive changes in crude oil price in the current period and one period prior are statistically significant to the world oil production. The second column shows that a negative change in oil price for the current period is statistically significant to BDI. In the third column, the estimated NARDL for crude oil price shows that the partial negative changes in oil production in the current period and one period prior, and negative change in BDI in the current period are statistically significant to the crude oil price. World oil production and BDI time trends are also statistically significant.

Table 3.8: NARDL estimation for world oil production, BDI, and crude oil price

	$lprod_t$	$lbdi_t$	lop_t
Coefficient estimates			
(ϕ_i)			
$lprod_{t-1}$	0.835*** (0.028)		
$lbdi_{t-1}$		1.026*** (0.050)	
$lbdi_{t-2}$		-0.127** (0.050)	
lop_{t-1}			1.393*** (0.050)
lop_{t-2}			-0.531*** (0.082)
lop_{t-3}			0.098** (0.049)
$(\beta_{j,i}^{+'}, \beta_{j,i}^{-'})$			
$lprod_t^+$		1.189 (1.580)	-0.073 (0.626)
$lprod_{t-1}^+$		-0.176 (1.528)	-0.375 (0.608)
$lprod_t^-$		0.160 (1.489)	-1.862*** (0.566)
$lprod_{t-1}^-$		0.312 (1.509)	1.469** (0.567)
$lbdi_t^+$	0.003 (0.006)		0.037 (0.038)
$lbdi_{t-1}^+$	-0.003 (0.008)		0.002 (0.055)
$lbdi_{t-2}^+$	0.004 (0.006)		-0.017 (0.039)
$lbdi_t^-$	0.002 (0.005)		0.069** (0.031)
$lbdi_{t-1}^-$	-0.003 (0.007)		-0.018 (0.046)
$lbdi_{t-2}^-$	0.004 (0.005)		-0.028 (0.032)
lop_t^+	-0.052*** (0.014)	0.144 (0.242)	
lop_{t-1}^+	0.047** (0.022)	0.023 (0.381)	
lop_{t-2}^+	0.017 (0.022)	0.031 (0.378)	
lop_{t-3}^+	-0.017 (0.013)	-0.145 (0.230)	
lop_t^-	0.006 (0.012)	0.467** (0.199)	
lop_{t-1}^-	-0.004 (0.020)	-0.156 (0.352)	
lop_{t-2}^-	0.011 (0.020)	-0.340 (0.351)	
lop_{t-3}^-	-0.009 (0.012)	0.014 (0.200)	
(a_1)			
t	3.9e-04*** (1e-04)	-0.005** (0.002)	8.6e-04 (6.4e-04)

Continued on next page

Table 3.8 – *Continued from previous page*

	$lprod_t$	$lbdi_t$	lop_t
(a_0)			
Constant	1.691*** (0.289)	2.195*** (0.668)	-0.110 (0.166)
Observations	418	418	418
F-statistic	3634.20	312.87	1473.58
Prob (F-statistic)	0.0000	0.0000	0.0000
Adjusted R^2	0.9929	0.9183	0.9802
Root MSE	0.0102	0.1774	0.0706

Notes:
Standard errors are in parentheses * $p < .10$, ** $p < .05$, *** $p < .01$

Table 3.9 presents NARDL estimates with a structural break in January 2009. The presence of the breaks causes a reduction in BDI and oil prices. The break in intercept (b_t) is statistically significant in BDI and oil price equations. The break interaction with BDI, both positive and negative shocks in slope ($b^*lbdi_t^+$ and $b^*lbdi_t^-$), is strongly significant in the oil price equation, whilst the break interaction with the negative shock of production ($b^*lprod_t^-$) is stronger than with the positive shock of production ($b^*lprod_t^+$). The BDI equation shows that the break interaction with positive and negative shocks in production is significant, while the break interaction with the prices ($b^*lop_t^+$ and $b^*lop_t^-$) is not. None of the breaks in the production equation is significant, and the magnitude is very small. The estimates with a break imply that crude oil price is affected much by the global demand that shocks both demand boom and drops and the disruption due to production decline. The production does not respond to the breaks because, in the short run, the elasticity of supply is inelastic. It takes a long time for production to respond to price and demand shocks.

Table 3.9: NARDL estimation with a structural break for world oil production, BDI, and crude oil price

	lprod_t	lbdi_t	lop_t
Coefficient estimates			
(ϕ_i)			
lprod_{t-1}	0.825*** (0.029)		
lbdi_{t-1}		0.983*** (0.051)	
lbdi_{t-2}		-0.132** (0.051)	
lop_{t-1}			1.331*** (0.049)
lop_{t-2}			-0.469*** (0.079)
lop_{t-3}			0.052 (0.048)
$(\beta^{+'}_{j,i}, \beta^{-'}_{j,i}, \gamma_0, \gamma^+_j, \gamma^-_j)$			
lprod^+_t		1.209 (1.565)	0.341 (0.606)
lprod^+_{t-1}		-0.117 (1.513)	-0.725 (0.587)
lprod^-_t		0.887 (1.492)	-1.685*** (0.546)
lprod^-_{t-1}		3.7e-04 (1.494)	1.576*** (0.547)
lbdi^+_t	0.001 (0.006)		0.058 (0.039)
lbdi^+_{t-1}	-0.003 (0.008)		-0.006 (0.052)
lbdi^+_{t-2}	0.005 (0.006)		0.008 (0.039)
lbdi^-_t	0.003 (0.005)		0.097*** (0.030)
lbdi^-_{t-1}	-0.004 (0.007)		-0.020 (0.044)
lbdi^-_{t-2}	0.004 (0.005)		-0.013 (0.032)
lop^+_t	-0.054*** (0.014)	0.088 (0.241)	
lop^+_{t-1}	0.047** (0.022)	0.022 (0.376)	
lop^+_{t-2}	0.017 (0.022)	-0.029 (0.372)	
lop^+_{t-3}	-0.015 (0.014)	0.035 (0.232)	
lop^-_t	0.010 (0.012)	0.484** (0.205)	
lop^-_{t-1}	-0.004 (0.020)	-0.068 (0.350)	
lop^-_{t-2}	0.013 (0.020)	-0.299 (0.347)	
lop^-_{t-3}	-0.012 (0.012)	-0.157 (0.207)	
b_t	0.038 (0.026)	34.960** (14.941)	14.722* (8.005)
b*lprod^+_t		-3.057** (1.334)	-1.238* (0.724)

Continued on next page

Table 3.9 – *Continued from previous page*

	lprod_t	lbdi_t	lop_t
b*lprod_t^-		-5.140** (2.122)	-2.325*** (0.649)
b*lbdi_t^+	-3.9e-04 (0.004)		-0.096*** (0.026)
b*lbdi_t^-	7.48e-04 (0.003)		-0.074*** (0.018)
b*lop_t^+	-0.008 (0.006)	-0.192 (0.126)	
b*lop_t^-	-0.008 (0.005)	-0.094 (0.105)	
(a_1) t	4.86e-04*** (1.47e-04)	-0.006** (0.003)	0.002** (8e-4)
(a_0) Constant	1.762*** (0.304)	3.005*** (0.828)	-0.147 (0.236)
Observations	417	417	417
F-statistic	2764.98	242.91	1191.20
Prob (F-statistic)	0.0000	0.0000	0.0000
Adjusted R^2	0.9929	0.9208	0.9819
Root MSE	0.0102	0.1747	0.0675

Notes:
Standard errors are in parentheses * $p < .10$, ** $p < .05$, *** $p < .01$

3.5.2.1 NARDL in error correction interpretation

Table 3.10 reports the results of the bounds test and presents the estimated NARDL in error correction form. In Panel (i) of Table 3.10, the bounds test indicates that the NARDL model also supports the presence of long-run asymmetry in world oil production and BDI equations because the F-statistic and t-statistic are large enough to reject the null hypothesis at a 5% significance level. The joint null hypothesis is that there is no long-run relationship in the asymmetry model. However, the long-run asymmetry is weak for the crude oil price equation. The F-stat and t-stat are within the inconclusive range of 10% critical values, but neither are large enough to reject the null hypothesis at a 5% significance level, and nor are they small enough not to reject the null hypothesis. The long-run relationship is present for the crude oil price because the long-run coefficients of partial positive and negative supply and demand changes are statistically significant. However, the similar magnitudes of the positive and negative supply and demand changes cause no asymmetry effect on the oil price.

Table 3.10: NARDL estimation in error-correction form for world oil production, BDI, and crude oil price

	D.lprod$_t$	D.lbdi$_t$	D.lop$_t$
(i) *Case 5*			
Bound Test H$_0$: no level relationship			
F-stat‡	7.150***	5.175**	3.658
t-stat‡	-5.914***	-4.718**	-3.663
(ii) *Adjustment factor (-α)*‡			
lprod$_{t-1}$	-0.166*** (0.028)		
lbdi$_{t-1}$		-0.101*** (0.021)	
lop$_{t-1}$			-0.040*** (0.011)
(iii) *Long-run* ($\theta_j^{+'}, \theta_j^{-'}$)			
lprod$^+_{t-1}$		10.020** (4.589)	-11.172** (5.039)
lprod$^-_{t-1}$		4.668 (5.203)	-9.825* (5.361)
lbdi$^+_{t-1}$	0.021** (0.008)		0.535** (0.213)
lbdi$^-_{t-1}$	0.020*** (0.007)		0.569*** (0.188)
lop$^+_{t-1}$	-0.029*** (0.011)	0.533** (0.266)	
lop$^-_{t-1}$	0.017 (0.016)	-0.142 (0.453)	
(iv) *short-run* ($\psi_{y,i}, \omega_j^{+'}, \omega_j^{-'}, \psi_{xj,i}^{+'}, \psi_{xj,i}^{-'}$)			
D.lbdi$_{t-1}$		0.127** (0.050)	
D.lop$_{t-1}$			0.433*** (0.049)
D.lop$_{t-2}$			-0.098** (0.049)
D.lprod^+_t		1.189 (1.580)	-0.073 (0.626)
D.lprod^-_t		0.160 (1.489)	-1.862*** (0.566)
D.lbdi^+_t	0.003 (0.006)		0.037 (0.038)
D.lbdi$^+_{t-1}$	-0.004 (0.006)		0.017 (0.038)
D.lbdi^-_t	0.002 (0.005)		0.069** (0.031)
D.lbdi$^-_{t-1}$	-0.004 (0.005)		0.028 (0.032)
D.lop^+_t	-0.052*** (0.014)	0.144 (0.242)	
D.lop$^+_{t-1}$	-2.6e-04 (0.014)	0.114 (0.240)	
D.lop$^+_{t-2}$	0.017 (0.013)	0.145 (0.230)	
D.lop^-_t	0.006 (0.012)	0.467** (0.199)	

Continued on next page

Table 3.10 – *Continued from previous page*

	D.lprod$_t$	D.lbdi$_t$	D.lop$_t$
D.lop$^-_{t-1}$	-0.001 (0.013)	0.326 (0.219)	
D.lop$^-_{t-2}$	0.009 (0.012)	-0.014 (0.200)	
(a_1)			
t	3.9e-04*** (1e-04)	-0.005** (0.002)	8.6e-04 (6e-04)
(a_0)			
Constant	1.691*** (0.289)	2.195*** (0.668)	-0.110 (0.166)

Notes:
Standard errors are in parentheses * $p < .10$, ** $p < .05$, *** $p < .01$
‡Lower-bound and upper bound critical values are based on Kripfganz et al. (2018) that covers a broader model specification and can be used for any number of short-run coefficients.

In Panel (iii) of Table 3.10, the long-run equilibrium relationship is presented for the three equations. In the first column of Panel (iii), the long-run equilibrium for world oil production indicates two main things. First, the coefficient is negative and significant for the positive oil shock, while being positive but not significant for the negative oil shock. Second, the estimated coefficients are very small; a 1% increase in oil price relates negatively with oil production by 0.03%, while a 1% decrease in oil price relates positively with oil production by 0.02% (which is not significant). In the second column of Panel (iii), the estimated long-run equilibrium for BDI shows two things. First, BDI indicates a positive and significant relationship with the positive oil shock. A 1% increase in oil price relates positively to the BDI by 0.53%. Second, BDI indicates a weak and insignificant relationship with the negative oil shock. A result similar to that of oil production equation is obtained, indicating that both supply and demand have a stronger link to positive oil price shock than to negative oil price shock.

In the third column of Panel (iii), the estimated long-run equilibrium for crude oil price indicates that the three equations generally report a positive relationship between oil price and oil demand, and a negative relationship between oil price and oil production, which is as expected. There are however two exceptions. First, a positive relationship between oil production and negative oil price shock is implied by the model of oil production, and second, a negative relationship between oil demand and negative oil price shock is implied by the model of oil demand. Nonetheless, these relationships are weak. Crude oil price indicates a stronger link with the positive and negative supply shocks than with the positive and negative demand shocks.

Turning to consistency of the magnitude of the estimated long-run relationship across the three equations, this is expressed in Equations 3.29-3.33 for the supply and oil price relationship, and Equations 3.34-3.38 for the demand and oil price relationship. These equations in the error correction form represent the cointegrating vectors on how the positive and negative shocks of supply, demand, and oil prices are related in the long run. To elaborate on the elasticity of price to demand and supply change comparison process, Equations 3.29-3.33 express the long-run relationship between oil price and supply. This is derived from three single equations: Equations 3.29 and 3.30 are based on the supply equation in second column, Panel (iii) of Table 3.10; Equations 3.31 and 3.32 on the demand equation derived from the third column, Panel (iii); and Equation 3.33 is based on the price equation in fourth column, Panel (iii).

The interpretation of the coefficient estimates in Equations 3.29-3.30 represent how supply relates to the positive and negative oil price shocks in the long run. Similar to the magnitudes in the long-run relationship of symmetric ARDL earlier, production relates to the positive and negative oil price shocks in a small magnitude in terms of percentage, as shown in the first line of Equation 3.29 (0.029 and 0.017, respectively). The sample variance is small relative to the sample mean for partial positive and negative supply shocks. Thus, a 1% response can be greater in terms of the sample. By way of example, the consistency of the magnitude of the estimated long-run relationship is checked by comparing the coefficient of -33.973 in Equation 3.29 (positive supply) with the coefficient of -11.172 in Equation 3.33 (price). Price impact's magnitude is thus three times that of positive supply. By then comparing the other coefficients between the equations, it may be identified which of the three gives the largest magnitude, which offers a consistency check.

$$
\begin{aligned}
lprod_{t-1} &= -0.029\, lop^{+}_{t-1} + 0.017\, lop^{-}_{t-1} \\
\Longleftrightarrow\; lop^{+}_{t-1} &= -33.973\, lprod_{t-1} + 0.586\, lop^{-}_{t-1} && (3.29) \\
\Longleftrightarrow\; lop^{-}_{t-1} &= 59.553\, lprod_{t-1} + 1.706\, lop^{+}_{t-1} && (3.30)
\end{aligned}
$$

$$
\begin{aligned}
10.020\, lprod^{+}_{t-1} + 4.668\, lprod^{-}_{t-1} &= -0.533\, lop^{+}_{t-1} + 0.142\, lop^{-}_{t-1} \\
\Longleftrightarrow\; lop^{+}_{t-1} &= -18.815\, lprod^{+}_{t-1} - 8.765\, lprod^{-}_{t-1} + 0.266\, lop^{-}_{t-1} && (3.31) \\
\Longleftrightarrow\; lop^{-}_{t-1} &= 70.466\, lprod^{+}_{t-1} + 32.828\, lprod^{-}_{t-1} + 3.754\, lop^{+}_{t-1} && (3.32)
\end{aligned}
$$

$$
lop_{t-1} = -11.172\, lprod^{+}_{t-1} - 9.825\, lprod^{-}_{t-1} \qquad (3.33)
$$

$$
\begin{aligned}
0.021\, lbdi^{+}_{t-1} + 0.020\, lbdi^{-}_{t-1} &= 0.029\, lop^{+}_{t-1} - 0.017\, lop^{-}_{t-1} \\
\Longleftrightarrow\; lop^{+}_{t-1} &= 0.729\, lbdi^{+}_{t-1} + 0.677\, lbdi^{-}_{t-1} + 0.586\, lop^{-}_{t-1} && (3.34) \\
\Longleftrightarrow\; lop^{-}_{t-1} &= -1.278\, lbdi^{+}_{t-1} - 1.187\, lbdi^{-}_{t-1} + 1.706\, lop^{+}_{t-1} && (3.35)
\end{aligned}
$$

$$
\begin{aligned}
lbdi_{t-1} &= 0.533\, lop^{+}_{t-1} - 0.142\, lop^{-}_{t-1} \\
\Longleftrightarrow\; lop^{+}_{t-1} &= 1.878\, lbdi_{t-1} + 0.266\, lop^{-}_{t-1} && (3.36) \\
\Longleftrightarrow\; lop^{-}_{t-1} &= -7.033\, lbdi_{t-1} + 3.754\, lop^{+}_{t-1} && (3.37)
\end{aligned}
$$

$$
lop_{t-1} = 0.535\, lbdi^{+}_{t-1} + 0.569\, lbdi^{-}_{t-1} \qquad (3.38)
$$

Likewise, Equations 3.34–3.38 express the long-run relationship between oil price and demand, which comes from three single equations: Equations 3.34 and 3.35 are based on the supply equation in second column, Panel (iii) of Table 3.10; Equations 3.36 and 3.37 on the demand equation derived from BDI equation in third column, Panel (iii); and Equation 3.38 on the price equation. Comparing, for example, the coefficient of 1.878 in Equation 3.36 (positive demand) to the coefficient of 0.535 in Equation 3.38 (price impact), then the impact of price is 3.5 times larger in magnitude than that of positive demand. By comparing the other coefficients between the equations, it may be identified, as a consistency check, which out of the three gives the largest magnitude.

Taking oil price as the dependent variable, the impacts on oil price of positive and negative demand shocks are greater than those exerted by supply shocks. From Equations 3.37 and 3.38 it may be seen that oil price needs to increase 13 times per unit of negative change in BDI. The coefficient is negative for the positive and negative supply shocks on price for all three equations (refer to Equations 3.29 and 3.31). The sign for the relationship that positive and negative supply shocks have on negative change in oil price is positive (refer to Equations 3.30 and 3.32), implying that a decrease in oil price relates positively to supply shock, which also reflects a negative relationship. The coefficient is positive for positive and negative demand shocks on price (refer to Equations 3.34 and 3.36). Likewise, a negative sign for the relationship that positive and negative demand shocks have on negative change in price means that a decrease in oil price negatively relates to demand shock. These signs do not represent causality, but the interaction across positive and negative shocks when they deviate from long run equilibrium.

Panel (iv) of Table 3.10 shows the short-run dynamics of the three equations. The first column indicates that the coefficients of demand in the supply equation are small. It implies that supply exhibits a limited contemporaneous response to demand shift. The coefficients of oil price in the supply equation are also small. Supply responds negatively to the positive oil price shock by 0.05% for a 1% increase in oil price. In contrast, in the second column, large coefficients of supply in the demand equation imply that demand responds rapidly (contemporaneously) to the supply shock, and this is more so for a positive supply shock than for a negative supply shock. Global demand responds rapidly to a negative oil price shock. A 1% decrease in oil price increases global demand by 0.47%. In the third column, the impacts of negative shocks in supply and demand on the oil price are significant. Oil price responds more rapidly to supply and demand disruptions than to positive supply and demand shocks.

The adjustment factors are reported in Panel (ii) of Table 3.10. Adjustment factors for the three equations are all negative and statistically significant, which means that the estimated relationships are stable in their error correction forms. When the world oil production is 1% higher than the long-run equilibrium level, it decreases by 0.17% in the subsequent period. If the BDI is 1% higher than the long-run equilibrium level, it decreases by 0.10% in the subsequent period. Finally, when the oil price is 1% higher than the long-run equilibrium level, it declines by 0.04% in the subsequent period.

The sign and magnitude of the three equations is examined for consistency as illustrated in Figure 3.4-3.6. The three equations have consistent signs but differ in their magnitudes depending on the partial positive or negative changes. Crude oil price relates negatively with oil production and positively with BDI, and oil production relates positively with BDI. Figure 3.4 shows that a larger magnitude of price adjustment is required for the positive supply shock (as shown by the solid line), and a larger supply adjustment is necessary to adjust to a positive price shock (as shown by the dashed line). A larger oil price adjustment is also required to adjust to a negative demand shock (as shown by the solid line) and a larger demand adjustment is needed for the positive oil price shocks (as shown by the dashed line), as illustrated in Figure 3.5. Figure 3.6 shows that supply needs a similar adjustment to positive and negative demand shocks (as shown by the solid line), and that demand requires a larger adjustment for the positive supply shock (as shown by the dashed line).

Table 3.11 presents NARDL model with a structural break in error cor-

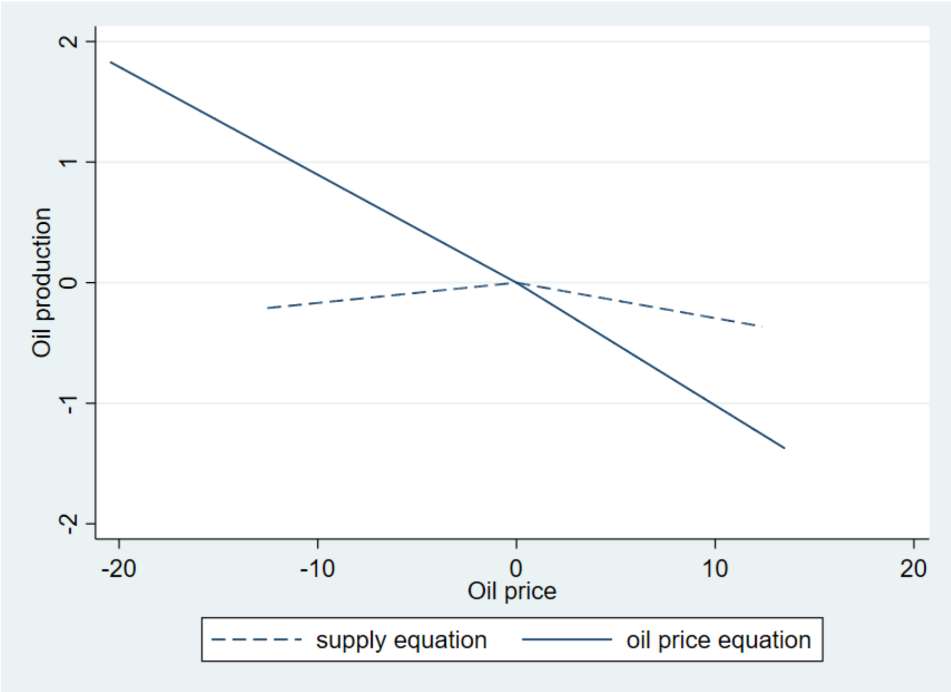


Figure 3.4: Oil price and supply adjustment

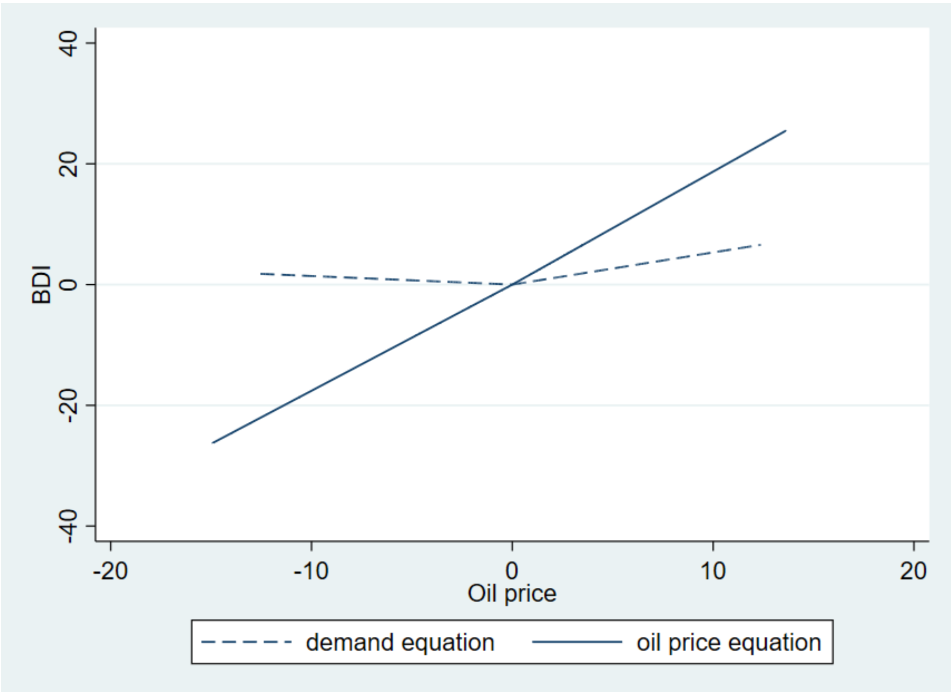


Figure 3.5: Oil price and demand adjustment

rection form. The model with a break in January 2009 shows the presence of the long-run equilibrium relationship as indicated by the rejection of the

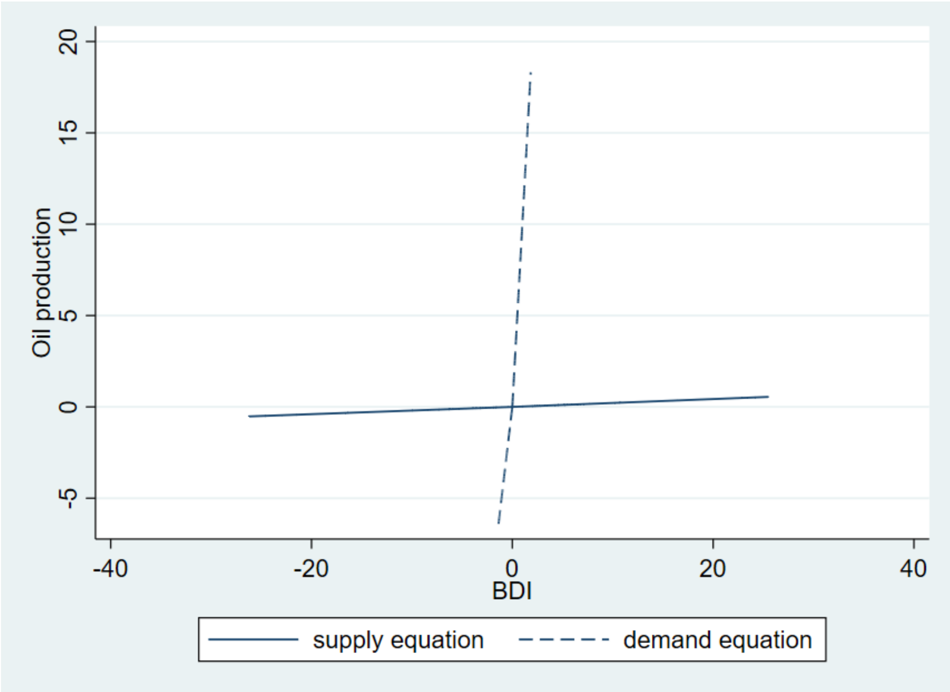


Figure 3.6: Oil supply and demand adjustment

Bounds test at 1% significance level for all three equations in Panel (i) of Table 3.11. The NARDL results confirm consistency with NARDL without a break in terms of the sign of the coefficients. The adjustment factor ($-\alpha$) in Panel (ii) is negative and statistically significant for all equations, which indicates the stable relationship for the dependent variable to respond when the three variables deviate from long-run equilibrium. The long-run coefficients (θ) also have consistent signs with the model without a break. In the long run, both positive and negative demand shocks are significant on the crude oil price. Their magnitudes are nearly similar, causing no dominant effect on oil prices by either positive or negative demand shocks. In the long run, the positive shocks of supply and oil prices are significant on demand.

Table 3.11: NARDL estimation with a structural break in error-correction form for world oil production, BDI, and crude oil price

	D.lprod$_t$	D.lbdi$_t$	D.lop$_t$
(i) *Case 5*			
Bound Test H_0: no level relationship			
F-stat‡	7.399***	7.189***	9.744***
t-stat‡	-5.978***	-5.678***	-5.976***
(ii) *Adjustment factor (-α)*‡			
lprod$_{t-1}$	-0.175*** (0.030)		
lbdi$_{t-1}$		-0.150*** (0.028)	
lop$_{t-1}$			-0.085*** (0.014)
(iii) *Long-run* ($\theta_j^{+'}, \theta_j^{-'}$)			
lprod$^+_{t-1}$		7.303** (3.445)	-4.505* (2.381)
lprod$^-_{t-1}$		5.939 (3.891)	-1.275 (2.463)
lbdi$^+_{t-1}$	0.021* (0.012)		0.698*** (0.112)
lbdi$^-_{t-1}$	0.015 (0.012)		0.751*** (0.196)
lop$^+_{t-1}$	-0.029 (0.018)	0.780*** (0.226)	
lop$^-_{t-1}$	0.038 (0.024)	-0.257 (0.574)	
(iv) *short-run* ($\psi_{y,i}, \omega_j^{+'}, \omega_j^{-'}, \psi_{xj,i}^{+'}, \psi_{xj,i}^{-'}, \gamma_0, \gamma_j^+, \gamma_j^-$)			
D.lbdi$_{t-1}$		0.132** (0.051)	
D.lop$_{t-1}$			0.417*** (0.047)
D.lop$_{t-2}$			-0.052 (0.048)
D.lprod^+_t		1.209 (1.565)	0.341 (0.606)
D.lprod^-_t		0.887 (1.492)	-1.685*** (0.546)
D.lbdi^+_t	0.001 (0.006)		0.058 (0.039)
D.lbdi$^+_{t-1}$	-0.005 (0.006)		-0.008 (0.039)
D.lbdi^-_t	0.003 (0.005)		0.097*** (0.030)
D.lbdi$^-_{t-1}$	-0.004 (0.005)		0.013 (0.032)
D.lop^+_t	-0.054*** (0.014)	0.088 (0.241)	
D.lop$^+_{t-1}$	-0.002 (0.014)	-0.006 (0.240)	
D.lop$^+_{t-2}$	0.015 (0.014)	-0.035 (0.232)	
D.lop^-_t	0.010 (0.012)	0.484** (0.205)	

Continued on next page

Table 3.11 – *Continued from previous page*

	D.lprod$_t$	D.lbdi$_t$	D.lop$_t$
D.lop$^{-}_{t-1}$	-5.17e-04 (0.013)	0.455** (0.219)	
D.lop$^{-}_{t-2}$	0.012 (0.012)	0.157 (0.207)	
b$_t$	0.038 (0.026)	34.960** (14.941)	14.722* (8.005)
b*lprod$^{+}_{t}$		-3.057** (1.334)	-1.238* (0.724)
b*lprod$^{-}_{t}$		-5.140** (2.122)	-2.325*** (0.649)
b*lbdi$^{+}_{t}$	-3.9e-04 (0.004)		-0.096*** (0.026)
b*lbdi$^{-}_{t}$	7.48e-04 (0.003)		-0.074*** (0.018)
b*lop$^{+}_{t}$	-0.008 (0.006)	-0.192 (0.126)	
b*lop$^{-}_{t}$	-0.008 (0.005)	-0.094* (0.105)	
(a_1) t	4.86e-04*** (1.47e-04)	-0.006** (0.003)	0.002** (8e-04)
(a_0) Constant	1.762*** (0.304)	3.005*** (0.828)	-0.147 (0.236)

Notes:
Standard errors are in parentheses * $p < .10$, ** $p < .05$, *** $p < .01$
‡Lower-bound and upper bound critical values are based on Kripfganz et al. (2018) that covers a broader model specification and can be used for any number of short-run coefficients.

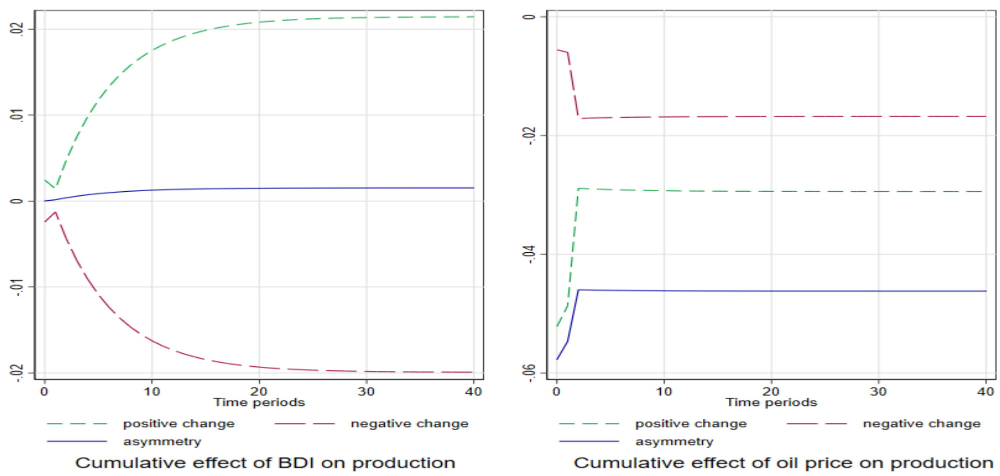
Figure 3.7 illustrates the dynamic multipliers of oil supply, global demand, and crude oil price. The blue asymmetry line indicates the cumulative impact of the partial positive and negative changes. The green and red lines respectively represent the cumulative impacts of a positive and negative unit change. In the figure, the red line refers to a long-run decrease (negative value of the long-run negative effect) for a more straightforward interpretation of the asymmetric effects.

The asymmetry line in Figure 3.7a shows that both positive and negative oil price changes following supply reduction have a greater magnitude for positive change than for negative. The oil production impacts appear to be greater after a positive change in demand in the short run. However, the accumulated impacts of positive and negative changes in demand do not substantially differ in the long run. The similar magnitudes cause the overall effect of cumulative global demand on world oil production to be more symmetrical. Figure 3.7b illustrates that positive supply shock has a dynamic impact on demand that is greater than the impact of a negative supply shock. However, the effect of the oil price shock on global demand shows that a positive oil price shock has a positive impact on demand. In contrast, a negative oil price shock initially has a negative effect on demand, but the effect turns positive after 16 periods. Figure 3.7c shows that negative supply change has a greater positive impact on oil price than that exerted by positive supply change in the short-run (i.e., up to around 20 periods). However, in the long run, a positive supply shock has a negative impact on oil prices that is greater than the positive impact of a negative supply change. The oil price impacts appear to be greater for negative demand change than for positive demand change in the short run. However, there is no substantial difference in the accumulated impacts of positive and negative demand change in the long run.

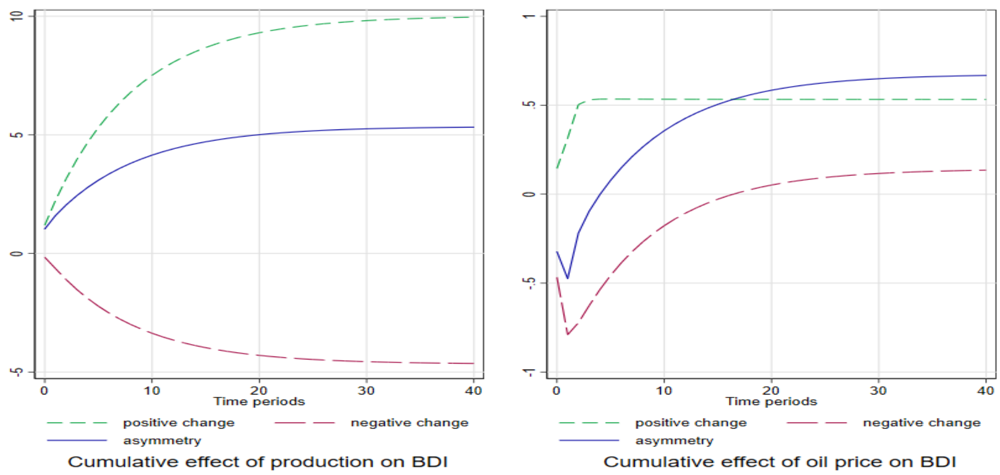
Figure 3.8 illustrates the dynamic multipliers of oil supply, global demand, and crude oil price for the NARDL model with a break in January 2009. Comparing it with the model without a break, the asymmetric effects in the model with a structural break are larger, particularly the positive shocks of oil price and production on global demand and the positive supply shock on the oil price. The positive and negative global demand shocks are significant but are symmetric.

The asymmetry line in Figure 3.8a shows that positive demand effects on supply are larger than the negative demand changes in the short run. In the long run, the accumulated impacts of positive changes in demand appear to be larger than the negative demand shocks on supply. The negative oil price changes have a slightly larger magnitude on supply than the positive shock but are overall insignificant. Figure 3.8b illustrates that a positive supply shock impacts demand more than a negative supply shock. The effect of the oil price shock on global demand shows that a positive oil price shock has a positive impact on demand. In contrast, a negative oil price shock initially has a negative effect on demand, but the effect turns positive after ten periods. The overall asymmetric effects of oil price on demand are negative for the first five months, then turn positive. Figure 3.8c shows that positive supply change has a greater negative impact on oil price than negative supply change in the short-run (i.e., up to around 20 periods). The oil price impacts appear slightly greater for negative demand change than for positive demand change in the short run. However, there is no substantial difference in the accumulated effects of positive and negative demand change in the long run.

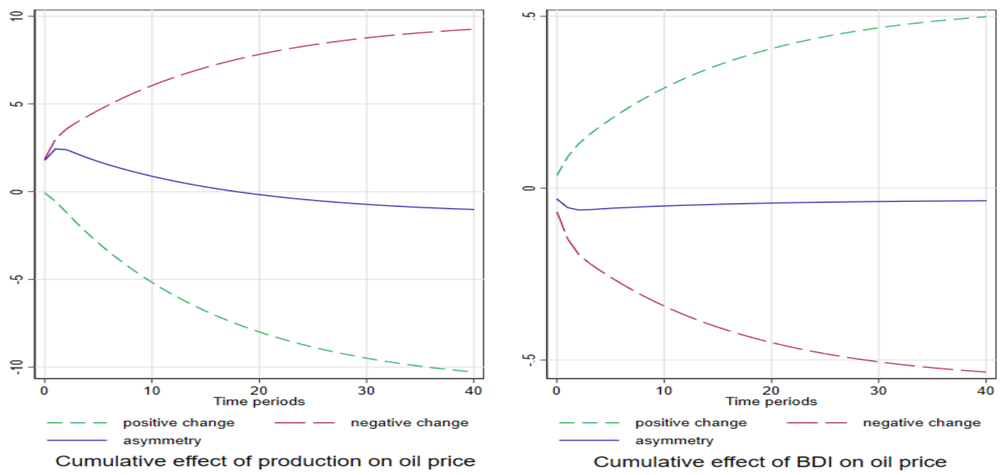
To sum up, having taken into account the joint null hypothesis of no long-run asymmetry relationship, the crude oil price exerts a statistically significant long-run asymmetrical effect on world oil supply, which is mainly caused by the positive shock of oil price. The fact that the effect is only seen in the long run is plausible as it is difficult to adjust production volume in the short run in response to sudden fluctuations in the global oil market. Crude oil price also has an asymmetric effect on global demand, with a significant impact exerted by positive oil price shock. World oil supply, and in particular partial positive changes to this, has significant asymmetric effects on global demand in the long run. Meanwhile, the long-run effects of global demand shock on the world oil supply have a more symmetrical effect. Global oil supply and demand also have symmetric effects on crude oil price in the long run.



(a) Asymmetric effects of global demand and oil price on world oil supply

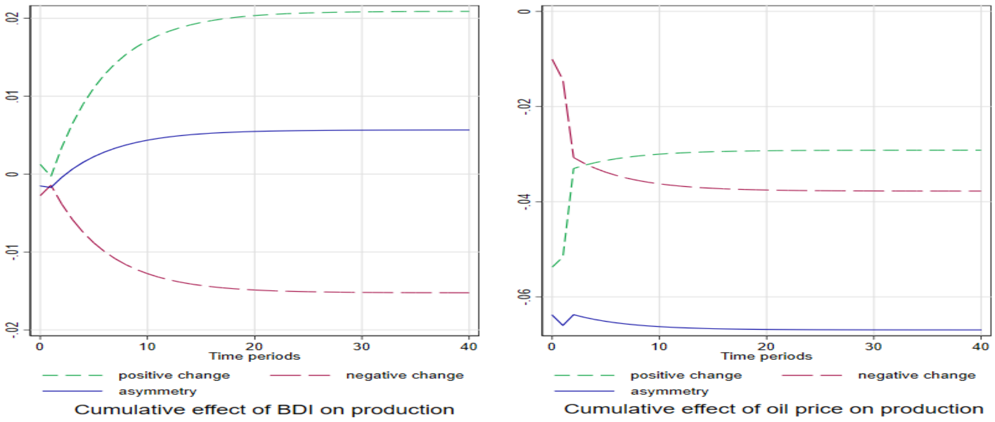


(b) Asymmetric effects of world oil supply and oil price on global demand

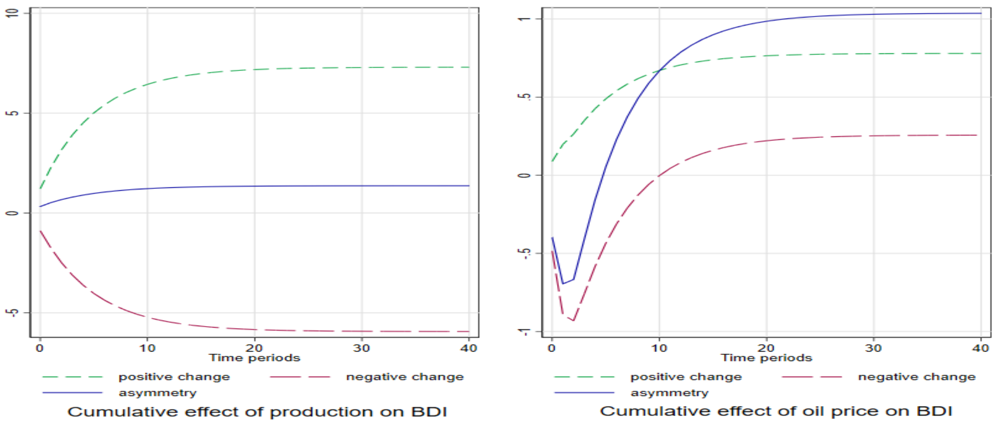


(c) Asymmetric effects of oil market variables on crude oil price

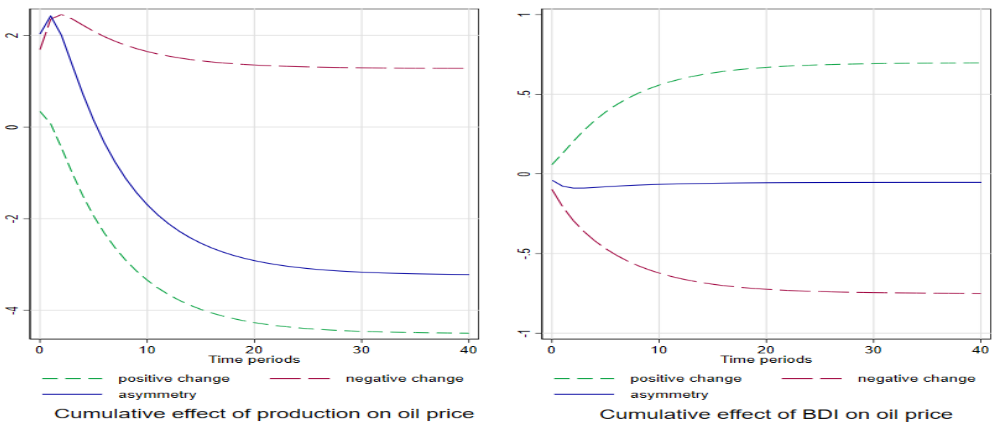
Figure 3.7: World oil supply, global demand, and oil price dynamic multipliers



(a) Asymmetric effects with a structural break of global demand and oil price on world oil supply



(b) Asymmetric effects with a structural break of world oil supply and oil price on global demand



(c) Asymmetric effects with a structural break of oil market variables on crude oil price

Figure 3.8: World oil supply, global demand, and oil price dynamic multipliers with a structural break

3.6 Conclusion

This chapter attempts to address two research questions: what is the relationship between crude oil supply, demand, and oil prices? And how do positive and negative shocks in the crude oil price affect oil supply and demand differently? In answering these research questions, this study contributes to the existing literature in three ways. First, the previous studies examine only short-run dynamics using first differences, whereas this study examines long-run relationships by applying the variables in levels. Second, the stationarity of the variables is carefully tested using a Fourier approximation-based unit root test that can accommodate the presence of structural breaks in variables. Most existing studies pay little attention to the stationarity of the variables in their application of the standard unit root test. This risks losing power when the series contains structural breaks. Furthermore, taking possible breaks in a time series into account is essential when choosing the appropriate empirical model. The unit root test indicates that crude oil price and BDI are non-stationary whereas world oil production is stationary. Thus, the ARDL model is applied to overcome the issue of mixed non-stationary and stationary variables. Third, this study examines the asymmetric effects in the relationship between global oil supply, demand, and crude oil price by applying the NARDL model.

The basic model of ARDL, interpreted with error correction, distinguishes long-run equilibrium from short-run dynamics relationships. The result finds a long-run relationship between global oil supply, demand, and crude oil price. In the long run, an increase in real oil prices is associated with a decline in world oil production and an increase in BDI. Furthermore, an increase in BDI is associated with increases in crude oil price and world oil production in the long run. When adjustment factors are compared across oil production, BDI, and oil price equations, it is evident that oil price must adjust more when the three variables deviate from the long-run equilibrium. In terms of short-run dynamics, the results imply that global oil supply and demand significantly affect crude oil price. Also that, oil prices significantly affect supply and demand, with a larger effect on demand in the short run.

The NARDL model accommodates the asymmetric effect by decomposing changes in the global oil production, BDI, and crude oil price into partial positive and negative changes. The NARDL error correction result finds a significant relationship among the three variables in the long run. Both supply and demand show a stronger link to a positive oil price shock than to a negative. Crude oil price relates negatively to the partial positive and negative supply shocks, and it relates positively to the partial positive and negative demand shocks. The oil price link to the positive supply shock is stronger than it is to the negative one, whilst its link to the positive demand shock is as strong as its link to the negative shock.

Turning to the relationship between supply and demand in the long-run equilibrium, supply links as strongly to the positive demand shock as it does to the negative one. The magnitude of the effect from a positive demand shock is similar to that from a negative demand shock, which means that the link between partial shocks is similarly balanced. Demand relates positively to positive and negative supply shocks, and it has a stronger link to a positive supply shock. The adjustment factor in the NARDL model requires more adjustment for demand when the three variables deviate from the long-run equilibrium. For the short-run dynamics of NARDL, the crude oil price impacts appear to be greater for demand changes that are negative than if they are positive. Demand impact is larger for negative oil price change. In contrast, supply impact is greater for positive oil price change, but is rather modest in magnitude.

The estimated asymmetric effects are consistent with the economic theory that an oil price increase is more significant to the economy than an oil price decrease. The empirical analysis in this chapter emphasises the supply-side role, which has been underestimated in most current literature. A long-run relationship between oil supply, demand, and oil price should be acknowledged, even if the crude oil price has little effect on world oil production in the short run. Analysing asymmetric effects is essential for understanding whether a positive or negative shock has a more dominant and significant impact on the global oil market in the short and long run. This analysis is beneficial because it allows industry to anticipate oil price fluctuation due to either a drop in the short-run global supply and demand or a surge in long-run supply. The asymmetric analysis allows oil producers and industrial manufacturers to initiate an appropriate response to the fluctuated oil price in the short run. In the long term and in response to the high oil price, the security of supply needs more attention to ensure that supply can meet global demand. In addition, the analysis is useful for assessing whether to immediately produce and sell more crude oil, or to hedge the price of crude oil production.

Two major events, COVID-19 and the Russia-Ukraine war happened in 2020 and 2022, respectively, after the thesis chapter was drafted. It would be interesting to examine their impacts on the global oil market for further research. Extending the observations to mid-2022 may have a few consequences, such as additional break dates in the temporal dynamics of each series and the relationship across oil production, demand, and oil price. Extending data sets may also affect the magnitudes of the asymmetric effects caused by many negative shocks throughout the observations. One of the potential challenges for such an extended analysis is that the break dates that occurred at the end of the sample period (e.g. Russia-Ukraine war) cannot be identified promptly due to the insufficient observations after a specified break date in the sub-sample. The researcher will have to wait for enough observations to be included in the last possible break date. In addition to further research, uncertainty is one of the factors that can affect the relationship between supply, demand, and price. Further analysis of uncertainty and the oil market is discussed in Chapter 4.

Chapter 4

Global oil market uncertainty, oil exploration, and crude oil price: an application of Google Trends

4.1 Introduction

Crude oil is critical to industry; thus, fluctuations in the crude oil price play a significant role in national and global economic growth and downturn. Over time, the behaviour pattern of the crude oil price is idiosyncratic. In the 1990s it was stable. It then began an upward trend from the turn of the century until 2008. From 2008 to 2010 there were extreme fluctuations caused by the Global Financial Crisis (GFC) and the Great Recession. In 2011, after these dramatic periods, crude oil prices remained stable until mid-2014, when there was another fluctuation episode which lasted until 2016. It then settled down until the end of 2019, when there was another period of fluctuation which has lasted to date. The most recent lowest price point for crude oil was in April 2020, when the WTI spot price was -37 USD per barrel and the Brent spot price was 9 USD per barrel.

A growing literature identifies the underlying causes of crude oil price behaviour. Kilian (2009) argues that the oil price surge has, as its underlying cause, the type of shock being experienced. Kilian (2009) also emphasises the importance of the global aggregate demand to the crude oil price. That is, demand corresponds to a strong economy. He also notes the impact of precautionary demand, which is the demand that refers to concerns about shortages in future supply. The types of shock and their consequential effects are not the same across the various episodes of turbulence. For instance, the increase in the crude oil price after 1979 was due to a combination of strong demand and a rise in precautionary demand. However, the oil price jump after 2003 was mainly driven by strong demand. Baumeister and Kilian (2016*b*) investigate the decline in the crude oil price from mid to end 2014 and conclude that there are predictable and unpredictable components causing the decline. The predictable components reflect the slowdown of the global economy and the expected oil production surplus. Meanwhile, the unpredictable components are associated with a combination of oil price expectation shock (via a drop in oil inventories) and the unpredictable economy. Hence, some crude oil fluctuation episodes are associated with uncertainty.

Given the role played by uncertainty in oil price volatility, the measures of uncertainty have received considerable attention in the recent literature. In a broader definition, the term 'uncertainty' refers to the unexpected outcome of various future circumstances that might apply to the economy, the financial markets, and geopolitical sectors. Three main approaches have been identified in the existing literature as measuring uncertainty as a transmission channel that affects crude oil price volatility. First, financial market-based uncertainty; second, macroeconomic uncertainty; and third, economic and policy uncertainty. However, there is no particular method and approach for measuring uncertainty as it specifically relates to oil demand and supply. The financial uncertainty typically approaches this by utilising a stock market volatility index as a proxy for measuring uncertainty, such as the crude oil volatility index (the so-called OVX index) which represents the near term option price of the crude oil as priced by the United States Oil Fund. In the current macroeconomic and political fields, current studies have two approaches for measuring uncertainty. The first is an indirect measure based on volatility in the forecasting of macroeconomic variables (Jurado et al., 2015; Rossi and Sekhposyan, 2015; Scotti, 2016) and the second is a direct measure of the public interest in discussions taking place in the media, particularly newspapers, reports, and the internet (Davis, 2016; Baker et al., 2016). Such public interest is measured by the frequency of keyword use.

In the global oil market, uncertainty reflects market concern about unexpected oil demand and supply; this concern affects the behaviour of the market, particularly in relation to the decision-making process about investments, such as oil exploration activity. In this study, the scope of uncertainty is not limited to speculation for the purpose of gaining future profit by crude oil hedging. It is more about unexpected circumstances in the short term, and the dramatic events that have attract public interest and consequences for the oil market. There is no exact procedure to measure the specific type of global oil market uncertainty that arises from supply and demand. It is therefore crucial to supplement the extant literature with an exploration of the interlinkage between oil market uncertainty and global oil market variables. From this, understanding can be gained, particularly for investors and decision-makers, about the uncertainty effects. Therefore, in a thorough examination of the global oil market framework, it is essential to take into account market concerns about supply and demand uncertainty and to examine how they affect the global oil market.

As the literature has not yet established an uncertainty measure specific to the oil market and its linkages to oil exploration, this study proposes to create a Google Trends-based uncertainty measure (i.e., the GTU index) for the global oil market. It analyses the effects of the GTU index on oil exploration and crude oil price. This study aims to answer three research questions. First, what difference might the construction of a more refined GTU index make to the literature and its findings? Second, what precisely do these GTU indices measure, and how do their measurements compare to those of the existing indices? Third, what is the relationship between uncertainty, crude oil price, and exploration?

Google Trends, a Google product, is one of the new direct measures of uncertainty. Google Trends captures, for a particular time and region, the number of searches made using a specific search term, which it can then rank according to its popularity. Now that internet technology is in the maturity stage, the broader public regard a web-based search engine as one of the most convenient ways of gathering news and updates. Google remains the most popular and widely used search engine. The notion of constructing an uncertainty index based on Google Trends is that the search index reflects public interest by measuring how frequently the public seek out information using this web-based search engine when there is uncertainty. The higher the uncertainty surrounding a particular topic or keyword, the more likely it is that the public will access the internet to search for it, and thus the frequency of the keyword will rank higher in Google Trends. Google trends thereby measures the direct observations of individuals' spontaneous behaviour towards uncertainty.

Google is the search engine platform that is, globally, most widely used. It is thus capable of capturing global public interest, and it generates a large number of observations. The data provided by Google Trends, as well as being directly observable, are free to access and publicly available. Furthermore, Google Trends has sufficient flexibility for it to be applied in various sectors, including the specific oil market field. Finally, GTU indices are, relatively speaking, easier to construct than the other uncertainty benchmarks that require extensive computational demand. It is therefore understandable that the literature on uncertainty measures has been extended from its traditional newspaper-base into an internet-based proxy, and also that the benefits and flexibility offered by the Google Trends-based approach have seen its utilisation as an index measuring uncertainty in the macroeconomic and financial market literatures for specific countries (for details, see Da et al. (2011); Dzielinski (2012); Castelnuovo and Tran (2017); Tran et al. (2019); Bilgin et al. (2019)).

Meanwhile, the current established literature focuses on five primary outcomes. First, there is a growing literature on the construction of macroeconomic, financial, and policy uncertainty measures. However, the literature developing an uncertainty proxy specifically for the global oil market remains scarce. To date, oil market uncertainty has been approached by looking at oil price volatility, utilising stock market indices and forecast-based oil prices. Second, the existing research in GTU has focused on crude oil price as a single component (see Qadan and Nama (2018); Li et al. (2019)) when constructing the basic GTU index. This approach is too restrictive to be an effective uncertainty measure as it neglects other related factors in the global oil market, most particularly supply and demand. Third, the focus of existing research mainly discusses the consequences of oil production, oil demand, and crude oil price shock on the uncertainty indices; it does not look at the effects of uncertainty on the global oil market. Only a small number of studies analyse the relationship between uncertainty and the global oil market (Kang and Ratti, 2013; Kang et al., 2017; Qadan and Nama, 2018). In assessing the global oil market, the extant studies approach uncertainty through a financial, macroeconomic, or policy uncertainty proxy instead of by identifying a specific proxy for the oil market that reflects oil supply and demand. Fourth, the existing studies analyse the uncertainty effects on the macroeconomy, stock market, and crude oil production. The effects of uncertainty on world oil exploration have not yet been explored. Fifth, the construction of uncertainty indices and their application have thus far been for a specific country (e.g. U.S., Australia, Italy, Turkey, and New Zealand).

As yet, no index of oil market uncertainty has had the capacity for worldwide application.

This research contributes to filling the gap in the existing studies in three ways. First, by its construction of a GTU that is a proxy measure of uncertainty specifically related to the global oil market. This research also extends the current literature by accommodating oil price as a single entity uncertainty measure and also by incorporating oil investment, oil supply, and oil demand as the main components of the measures in the proposed GTU indices. Second, after explaining how the indices are constructed, this study then compares the results obtained from the GTU indices to those of the existing uncertainty measures. From this, the study examines whether the GTU indices can offer a new approach for proxying global oil market uncertainty. Third, this study analyses the GTU oil market uncertainty effects on two specific oil market variables: oil exploration and crude oil price. This fills a gap not only in the oddly scant literature on oil market specific uncertainty but also in the literature on oil exploration. Furthermore, this research utilises worldwide data rather than data related to a specific country; it thus has greater generalisability than the findings from the current country-specific literature.

This study constructs the GTU oil market specific indices by extending the work of Guo and Ji (2013); Qadan and Nama (2018); Li et al. (2019) using the procedure proposed by Castelnuovo and Tran (2017). As well as the basic GTU index that incorporates a single oil price component (hereafter, 'GTU oil price'), there are two more refined types of GTU index proposed in this study. The first type relates to specific components. Thus, there are indices for GTU oil investment, GTU oil supply, and GTU oil demand, with each component being distinguished. The second type aggregates the three distinguished components into one comprehensive uncertain measure, termed the 'GTU oil market specific' index. Within the scanty oil market uncertainty literature, some studies construct GTU based on the basic index only, which means the uncertainty measure is restricted to only one component. While the basic index from Google Trends has already been applied as a proxy of oil market uncertainty (see (Qadan and Nama, 2018; Li et al., 2019)), these studies utilise either 'oil price' and its variant ('price of oil') or 'Brent' as the search term in Google Trends. This does not take into consideration other factors that might affect the global oil market. This approach is highly restrictive and limits the measure of global oil market uncertainty. Guo and Ji (2013) is the only paper that constructs GTU for oil demand. However, their choice of search terms does not actually reflect oil demand. For instance, their study utilises the phrase 'oil production' for the GTU oil demand component, which is inappropriate given that oil production reflects oil supply rather than oil demand.

This study applies Granger Causality and the Vector Autoregressive (VAR) framework to analyse the effect of the GTU index as a measure of uncertainty in oil exploration and the crude oil price in the short run. Crude oil price is one of the main indicators in the global oil market. The base model applies the U.S. real acquisition cost of imported crude oil price as the global oil price proxy, while the Brent real oil price is applied for the robustness test. Oil exploration is an important parameter for measuring investment activity in the oil market sector. Furthermore, and as argued in the well-established literature (Kellogg, 2014; Toews and Naumov, 2015), oil exploration significantly responds to crude oil price. This research will examine how oil market uncertainty affects global oil exploration with dominant non-conventional and conventional oil resources. Oil rig counts, representing global oil exploration, will be distinguished into North America (i.e., the U.S. and Canada) oil rig counts because of the shale oil boom in North America since the 2000s, and rest of the world rig counts (dominated by traditional oil exploration).

The findings of this research confirm that GTU indices hit high points at times when dramatic circumstances occurred; such occurrences would therefore be an interesting warning indicator. The uncertainty shocks are evident in the public reaction to three remarkable fluctuations in crude oil price over the past two decades. These followed the GFC and recession during 2008–09, an oil price drop in 2015–16, and the Coronavirus pandemic in 2020. The newly proposed GTU has positive and high correlation values with existing uncertainty benchmark indices; for example, the correlation with OVX oil volatility index is 0.83, while the correlation with macroeconomy uncertainty index proposed by Jurado et al. (2015), JMU, is 0.74. The positive and high correlation between GTU and the existing benchmark indices, particularly OVX and JMU, confirm the ability of the GTU index to capture dramatic events that lead to uncertainty shocks.

The unidirectional Granger causality from uncertainty to crude oil price is found in most of the uncertainty indices. This study also finds causality from most of the uncertainty indices to world oil exploration, but not for the North American oil exploration. This finding can be seen as conventional world oil exploration being more affected than non-conventional exploration by uncertainty. The tight oil exploration activity typical of the U.S. tends to exploit the big size of oil fields at the beginning of the boom era. Exploration activity declines over time and is more focused on continuing with existing projects rather on making new discoveries. Hence, uncertainty does not affect new exploration activity much in North America. The causality from crude oil price to oil exploration in all uncertainty models is expected. Exploration responds more than production to crude oil price fluctuation in the short run (Toews and Naumov, 2015; Kilian, 2009; Kilian and Murphy, 2014).

The impulse response functions show that uncertainty affects oil exploration and crude oil price. The unanticipated increase in GTU index causes a significant decline in world oil exploration, but it is less significant for North American oil exploration. The crude oil price shocks have a significant and positive response on world and North America oil exploration. The unexpected rise in GTU index also causes a negative response on the crude oil price. Compared to the benchmark uncertainty indices, the newly proposed GTU indices better explain the variability in oil exploration and crude oil price. Overall, uncertainty shocks contribute more to variability in world oil exploration than to variability in North America oil exploration. GTU oil supply contributes 14% to world oil exploration while the highest contribution to the North America oil exploration is made by GTU oil market specific, and that is only 6%.

This chapter is organised as follows. Section 4.2 discusses the literature on the well-established uncertainty indicators, which are set as the benchmark. Section 4.3 describes the existing uncertainty benchmark measures and global oil market data. Section 4.4 goes into detail about how the proposed GTU indices measure uncertainty. Section 4.5 presents the VAR framework. Section 4.6 presents empirical analysis on the relationship between uncertainty, world and North American oil exploration, and crude oil price. Section 4.7 concludes the chapter and suggests potential avenues for future research.

4.2 The literature on uncertainty measures and contribution of the study

The impact uncertainty has on various economic, financial, and geopolitical activities has prompted the academic literature to focus on how to measure uncertainty. However, there is no universally accepted methodology for quantifying uncertainty. The well-established literature measures uncertainty in a variety of ways based on the financial markets, forecasting, and public interest.

The financial market-based uncertainty measures apply various options, futures contracts, and other asset products to the exchange market. Forecast-based measures include estimating forecast error, standard deviation, or the variance of the variable of interest (see, for instance, Elder and Serletis (2010); Bachmann et al. (2013); Jurado et al. (2015); Rossi and Sekhposyan (2015); Jin (2019); Ahmadi et al. (2019)). Finally, the public interest measure relies on capturing the trends in behaviours reflecting market concern, which includes how frequently the public discuss and react to certain topics reported by the media. The source of data for this measure might be a country report, or newspaper-based (Baker et al., 2016; Davis, 2016; Ahir et al., 2018; Caldara and Iacoviello, 2018), or internet-based (Dzielinski, 2012; Guo and Ji, 2013; Ji and Guo, 2015; Bontempi et al., 2016; Castelnuovo and Tran, 2017; Qadan and Nama, 2018; Li et al., 2019; Tran et al., 2019; Shields and Tran, 2019). The main notion is that the discussion of certain popular topics attracts attention from the public, leading them to seek out information about those topics. A higher public tendency to access such information reflects a higher degree of uncertainty.

Returning to the literatures on the financial uncertainty measures, the most common such measure is the stock market index. One of the typical indices commonly used, particularly in the oil market, is the crude oil price volatility

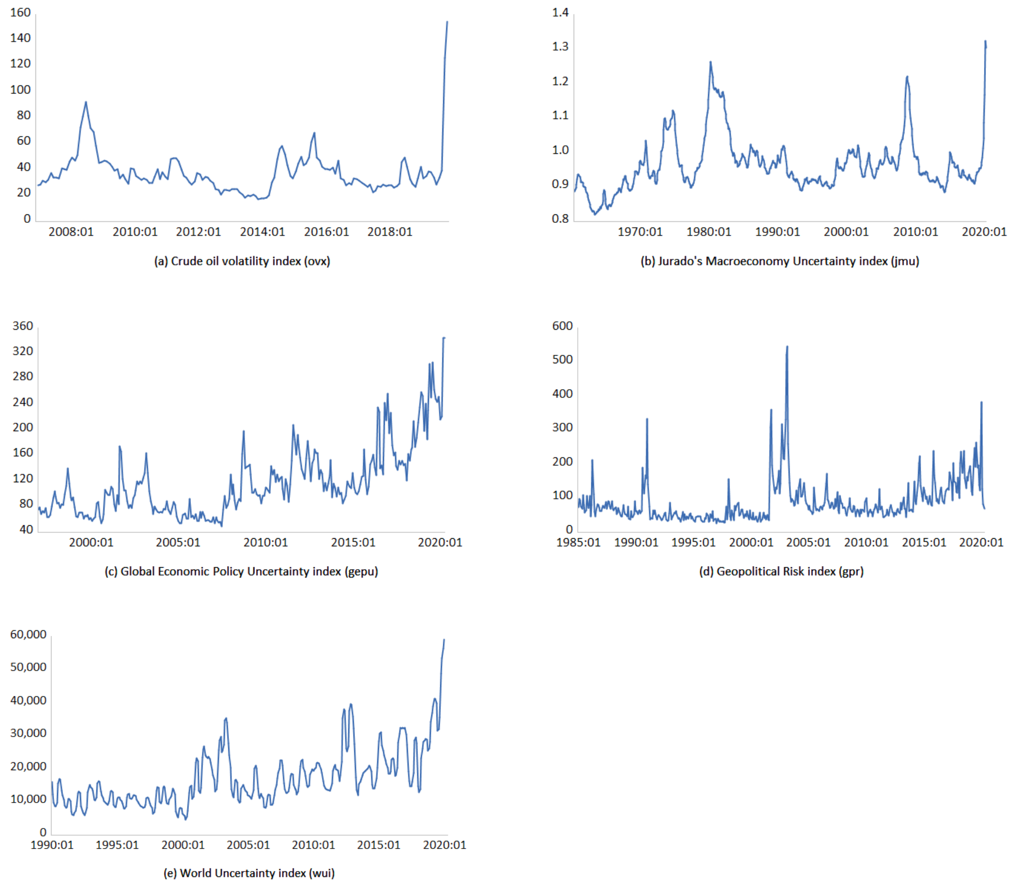


Figure 4.1: The Uncertainty Benchmark Measures for full sample period

OVX index. The OVX index measures short-term (i.e., 30-day) volatility in the market's expectation of the WTI price traded in the United States Oil Fund (CBOE, 2020). The oil volatility index has the advantage of measuring market behaviour as it relates to investment portfolios; thus, it can capture periods when investors are uncertain about the oil price. However, not all the volatility in the stock market index is a reflection of uncertainty. The fluctuation of the index can be associated with the heterogeneity of the risk and return assessment of the asset portfolio. Figure 4.1 shows that significant spikes in the OVX index may be seen in December 2008, February 2015, February 2016, and April 2020. The graph illustrates the high financial uncertainty occurring during the Great Recession of 2007–09, the crude oil price downturn in early 2015 and 2016, and the Coronavirus pandemic peak in April 2020. The graph also indicates

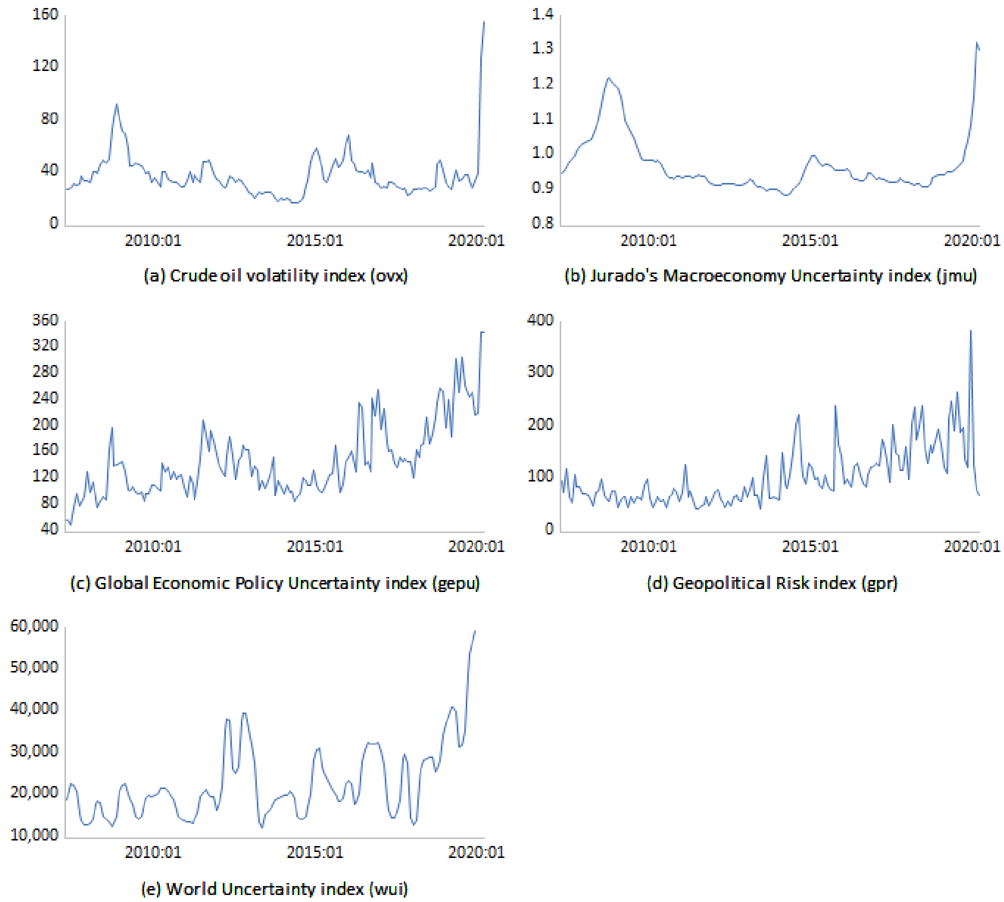


Figure 4.2: The Uncertainty benchmark measures for sample period: May 2007-April 2020

that the financial uncertainty index reaches a peak when the crude oil price fluctuates either in the high or the low state.

Macroeconomic uncertainty has been widely discussed in the literature through both direct and indirect measures of macroeconomic variables. The direct measure reflects the directly observable uncertainty events perceived by the public; for instance, via newspaper or internet-based uncertainty indices. The country report or newspaper-based indices focus more on how the media or journalists interpret and write the news, whereas the internet-based index focuses on the individual actions of the news recipients, which reflect how they perceive the news (Bontempi et al., 2016). As regards the indirect measures of uncertainty, these are derived from a latent process that is independent of any observable economic indicators (Bontempi et al., 2016). For example, Jurado et al. (2015) construct their macroeconomic uncertainty index based on the conditional volatility of the forecast value of various macroeconomic categories.

This particular forecast-based method requires the input of large datasets and it has the advantage of reducing bias in the sampling. However, the forecast-based method is not directly linked to uncertainty as a function of public interest. Furthermore, it requires intensive computational resources to treat the data (Bontempi et al., 2016). Compared to the OVX index, Jurado's macroeconomic uncertainty index produces fewer and less frequent spikes and it is less volatile, but it tends to capture business cycle episodes with larger magnitudes and more persistent effects. As shown by Figure 4.1, the macroeconomic uncertainty index indicates four significant spikes in December 1974, April 1980, October 2008, and April 2020. These peaks are associated with the 1973–75 recession, the early 1980s recession, the Great Recession of 2007–09, and the Coronavirus pandemic peak in April 2020. Like financial market uncertainty, macroeconomic uncertainty reaches high points when crude oil prices are in a high or low state.

The policy uncertainty index has been proposed by Baker et al. (2016) and extended by Davis (2016) via a newspaper-based direct measure. The index is made up of the relative frequency of three words related to economic, policy, and uncertainty terms in newspapers of 21 countries; the chosen countries are those that contribute for two-thirds of global output. For the global measure, Davis (2016) extends the work of Baker et al. (2016) by weighting the GDP average of each country. The newspaper-based uncertainty index has the benefit of being a directly observable measure of public reaction to a particular topic or event covered by the media. However, it requires the researcher to have access to the newspaper archives and to use a specific tool to count the words. Compared to the other uncertainty indices, as shown in Figure 4.1, the global economic policy uncertainty index exhibits many peaks with upward trends in September 1998, September 2001, March 2003, October 2008, August 2011, September 2015, June 2016, January 2017, December 2018, August 2019, and April 2020. These peaks correspond respectively with the Asian Financial Crisis, the 9/11 attack, Iraq invasion by the U.S., the Great Recession, the Eurozone crisis of mid-2011 to early 2013, the Chinese stock market bubble of mid-2015 to early 2016, the Brexit referendum, U.S President Trump's election, the U.S dollar debt risk to emerging markets (Kupelian and Jakeman, 2018), the U.S-China trade tension (IMF, 2019), and the Coronavirus pandemic.

Other uncertainty measures include geopolitical and world uncertainty indices. The geopolitical risk index proposed by Caldara and Iacoviello (2018) is a newspaper-based index focusing on global tensions, including military, war, threats, and nuclear incidents. As shown in Figure 4.1, the significant peaks occurred in April 1986, January 1991, October 2001, March 2003, and January 2020. These are associated with the nuclear accident at the Chernobyl plant, the Gulf War of 1990–91, the 9/11 attack, the Iraq invasion of 2003, and the U.S-Iran tension of 2020. The world uncertainty index proposed by Ahir et al. (2018) is a country report-based index measuring the general term of uncertainty as captured by the Economist Intelligence Unit through the keyword 'uncertainty'. This index has fewer peaks, in terms of both magnitude and frequency, than the other uncertainty indices. These occur in May 2003, December 2012, and January 2020, corresponding with the SARS outbreak, the Iraq invasion, the eurozone debt crisis, and the early phase of the Coronavirus pandemic.

The financial, macroeconomic, and global economic policy uncertainty indices all capture two significant spikes of uncertainty: the Great Recession of 2007–09 and the Coronavirus pandemic peak in April 2020. The uncertainty indices are also high when crude oil prices are extremely high, such as during the oil crisis and the recession in 1973–74, the Iranian revolution in 1979, the

Iran-Iraq war and the recession in 1980, and the Great Recession of 2007–09. But they are also high when oil prices in low such as during the oil price drop in early 2015–16 and Coronavirus pandemic in early 2020.

The literature on oil price behaviour also emphasises some episodes when oil price fluctuation has much more significant effects on the economy compared to other episodes. Immense contributions made by oil price shocks to the U.S macroeconomy prior to the 2008 were seen during the oil crisis in 1973–74, the Iranian revolution in 1979, and the Gulf War in 1991 (Gronwald, 2008; Hamilton, 2011). This view indicates that uncertainty, oil prices, and the economy may reinforce each other. However, the financial market uncertainty index is the only index that captures uncertainty during the crude oil price downturn in early 2015–16. To align the indices with the specification of their measures, it can be noted that the geopolitical risk index captures political events and is thus not directly capable of capturing the significant breaks in economic-related events, while the world uncertainty index measures various forms of general uncertainty ranging from global outbreaks to domestic economic and political uncertainty.

Despite this array of uncertainty measures, there is no specific uncertainty measure for the global oil market. Some of the established literature defines uncertainty as related to market expectation, precautionary demand, and speculation on forward-looking price; examples of these include the work of Apergis and Miller (2009); Kilian (2009); Alquist and Kilian (2010); Kaufmann (2011), and their extensions (Baumeister and Peersman, 2013; Fattouh et al., 2013; Kilian and Murphy, 2014; Yin and Zhou, 2016; Jin, 2019). The expectation of the future market is associated with concern about the lack of future supply and its higher future price, which trigger the market to make immediate purchases of crude oil. This research believes that the spread of the future price can predict the current price, and views uncertainty as speculation to gain profit. Thus, the most common proxy used to measure oil price uncertainty is crude oil inventory, which refers to the crude oil stock above ground. The existing studies also approach oil market uncertainty through the variance of the oil price forecast, as proposed by Elder and Serletis (2010). These authors measure oil price uncertainty as to the standard deviation of in the forecast error of oil price changes by applying the Generalised Autoregressive Conditional Heteroskedasticity (GARCH) framework. Baumeister and Peersman (2013), on the other hand, argue that uncertainty is derived from the greater oil price volatility that stimulates the crude oil options in the derivative market. Jo (2014) also models the oil price uncertainty through the volatility of the forecast error with but uses the stochastic volatility approach rather than the GARCH process. Kellogg (2014) applies future options of the oil price as a measure of oil price uncertainty.

The remaining uncertainty literature takes a very different approach to those outlined above for defining oil market uncertainty. It interprets uncertainty as public reaction, captured by the internet, to certain dramatic circumstances, which leads to oil price fluctuation. This measure derives a news-based direct measure of uncertainty from the widely used web-based search engine, Google. Google has created a product called Google Trends, which captures the popularity of certain keywords (hereafter referred to as search terms) relative to other 'Googled' search terms.

However, the extant implementations of Google Trends-based uncertainty (GTU) mostly measure macroeconomic uncertainty (Bontempi et al., 2016; Castelnuovo and Tran, 2017; Tran et al., 2019; Shields and Tran, 2019) and economic-financial uncertainty (Da et al., 2011; Dzielinski, 2012; Bilgin et al., 2019). For example, Da et al. (2011) use the GTU index to capture the stock market and conclude that an increase in GTU financial market uncertainty contributes to higher stock price. Dzielinski (2012) captures economic uncertainty based on GTU and finds a significant relationship with the stock market return. Bontempi et al. (2016) construct a GTU index for the macroeconomy from macroeconomy-related search terms, which is in line with previous work by Baker et al. (2016). The authors apply the VAR framework to examine the relationship between GTU index and U.S. macroeconomic variables and find that uncertainty significantly reduces output in the short term.

Castelnuovo and Tran (2017) construct the GTU macroeconomic index for the U.S and Australia. Their work confirms the positive correlation between their GTU index and the existing uncertainty measure for both countries. They also conclude that uncertainty causes a significant increase in the unemployment rate and a reduction in price. In an extension of the work of Castelnuovo and Tran (2017), Bilgin et al. (2019) develop GTU economic and financial indices for Turkey and conclude that an increase in GTU macroeconomic shocks decreases the stock market return and increases the interest rate and unemployment rate. Tran et al. (2019) and Shields and Tran (2019) extend the works of Bontempi et al. (2016) and Castelnuovo and Tran (2017) by investigating macroeconomic uncertainty at different regional levels. Shields and Tran (2019) use disaggregated data on U.S. state-level uncertainty and find a heterogeneous effect of uncertainty in the macroeconomy due to variations in fiscal policy and industry composition. Tran et al. (2019) propose a GTU index for the New Zealand macroeconomy and conclude that the GTU index is capable of predicting the GDP of New Zealand.

Meanwhile, the few contributions made by Google Trends-based indices to the extant oil market literature are limited to work by Guo and Ji (2013), Ji and Guo (2015), Qadan and Nama (2018), and Li et al. (2019). Current studies incorporate Google Trends to approach the crude oil market in two ways. First, by constructing a basic measure that has oil price as its only component. Second, by incorporating specific events that may lead to oil price fluctuation. Such studies use Google Trends to examine the effects of market concerns on the various oil price markets. For example, Guo and Ji (2013) apply oil price, oil demand, financial crisis, and the Libyan war as the market concerns captured by the GTU to measure short- and long-run relationships with the Brent price. They extend this research by investigating the oil market through oil-related event search terms. Thus, 'Hurricanes' are a proxy for supply shock, 'global financial crisis' proxies for aggregate demand shock, and the 'Libyan war' and 'OPEC conferences' proxy for precautionary demand (Ji and Guo, 2015). Qadan and Nama (2018) apply Google Trends for the search term 'oil price' and its variants, such as 'crude oil' and 'price of oil'. They regard this as an alternative investment sentiment indicator and examine its causality on oil returns. Li et al. (2019) use 'Brent' as the search term in Google Trends to measure public concerns about various oil price markets, namely the WTI, Brent, Dubai, and Daqing oil markets.

4.2.1 Uncertainty and crude oil price

The extant literature commonly applies the VAR framework or its extensions to address the following two types of research questions. The first analyses the predictability of uncertainty to oil market variables. Thus, Baumeister and Kilian (2016*b*) apply reduced form VAR to predict the cause of the oil price decline between mid to end 2014 and argue that inventories contribute to the unpredictable components of the decline. The second examines the relationship between uncertainty and other variables of interest by decomposing their variances into the shocks contributing to other variables. Such studies focus more on the effect of oil market shocks (e.g., oil price and oil production shocks) on uncertainty rather than on the consequences of uncertainty shocks for the global oil market. For example, Kang and Ratti (2013) find that an oil price increase due to precautionary demand is associated with an increase in the U.S. economy and policy uncertainty. However, a crude oil price increase due to global demand causes a decline in uncertainty.

Meanwhile, Degiannakis et al. (2018) do not support the argument that oil price shock leads to higher uncertainty, but rather argue that the uncertainty response to oil price shock is time-varying, having a short- to medium-run positive response at the beginning of the sample and a negative response in the latter part of sample. Furthermore, they find that the response of the uncertainty indices to oil price shock is heterogeneous. The positive response derives from macroeconomy uncertainty, the negative response derives from financial and commodity-related uncertainty, and economic and policy uncertainty has no response to oil price shock. They also argue that the relationship between uncertainty and oil price in the long term is insignificant.

A few of the contributions that examine the effect of uncertainty on the crude oil price find that the crude oil price responds negatively to uncertainty shocks in most of the sampled period, with some positive responses only in specific periods. For instance, Bekiros et al. (2015) confirm that economic policy uncertainty can forecast crude oil price changes. Li et al. (2019) investigate the various types of investor attention indices, including Google Trends, and find unidirectional Granger causality from investor attention to the WTI oil price. Kang and Ratti (2013) find that between eleven and fifteenth months after the shock, oil price has a significant negative response to economic and policy uncertainty. Aloui et al. (2016) also apply an economic policy uncertainty index and investigate its effect on the crude oil price return. They find that uncertainty has a positive effect on the crude oil return prior to the global financial crisis and Great Recession. However, over the entire sample, the crude oil return responds negatively to the uncertainty shock.

An increase in uncertainty triggers lower productivity and, in consequence, a lower demand for crude oil. Bloom (2009) uses the simulated method of moments model for firm-level data arguing that a higher uncertainty makes firms pause the hiring and investment activity temporarily, generating a rapid drop and rebound in output, employment, and productivity growth. Jurado et al. (2015) use hours worked and industrial production instead of labour productivity and find that a large macro uncertainty shock leads to a decline in production hours and employment. Macro uncertainty explains much larger during a recession than expansion. By applying forward refining margin volatility from future prices of commodities as the market uncertainty measure, Dunne and Mu (2010) find that an increase in market-level uncertainty reduces the probability of investment in the US oil refinery, which supports the finding of Dixit and Pindyck (1994) through their real-options approach that uncertainty causes the delay in the investment decision making. Empirically, they find that the probability of investment is reduced by 11% for one standard deviation of refining margin uncertainty.

The work of Antonakakis et al. (2014) supports the significant negative response of crude oil price to the economic policy uncertainty shock. Qadan and Nama (2018) apply some sentiment indicators, one of which is the sentiment index developed by Baker and Wurgler (2007). The other indicators are the financial stress index, OVX, and VIX. They find that the oil price response to these indicators is a significant decrease. Qadan and Nama (2018) also find that oil price responds temporarily and negatively to the Consumer Confidence Index, and positively to economic policy uncertainty and the consumer sentiment index.

Qadan and Nama (2018) confirm that there is a bidirectional Granger causality between GTU oil price for the U.S. and oil price returns. Guo and Ji (2013) find a significant short-run relationship between GTU for oil-related events, namely the financial crisis and the Libyan war, and the Brent oil price. Following their earlier study, Ji and Guo (2015) conclude that the uncertainty effects on crude oil price are contingent on the type of uncertainty event. Market concerns related to financial crisis and an increase in OPEC production, as measured by GTU, cause a significant negative response in oil price returns; meanwhile, GTU shocks due to hurricanes, the Libyan war, and OPEC production cuts lead the positive response on oil price returns volatility.

4.2.2 Uncertainty and economic activity

The existing literature mainly discusses the impact of uncertainty on investment as part of economic activity. There are two main conclusions from this discussion. First, as found by Antonakakis et al. (2016), uncertainty can predict investment. This study finds that economic policy uncertainty is able to predict sustainable returns to investment, particularly after the global financial crisis period. Second, there is a negative relationship between uncertainty and investment, with an increase in uncertainty triggering a decline in investment activities.

Most of the studies approach investment activities through macroeconomy indicators such as industrial production, GDP, expenditure, and employment rate. Sadorsky (1999) argues that oil price and oil price volatility play significant roles in economic activity (proxied by industrial production and GDP) between January 1986 and April 1996. Oil price volatility causes a decline in company earnings; this is associated with the role of oil as a production cost. Bloom (2009) applies stock market volatility index as the proxy of uncertainty and concludes that higher uncertainty causes firms to pause their investment activity. Jo (2014) concludes that oil price uncertainty causes a significant drop in real activity, which is proxied by industrial production. Jurado et al. (2015) find that their macroeconomic uncertainty index is correlated with real activity proxied by industrial production, working hours, and employment; they also find that an increase in the uncertainty index reduces real activity. Davis (2016)'s finding also supports the view that uncertainty shock decreases industrial production, which represents a country's economic activity.

Some of the studies consider expenditure decrease to be a measure of economic activity slowdown. Elder and Serletis (2010) look at crude oil forecasting dispersion and conclude that higher uncertainty in oil prices leads to a steep decrease in mining expenditure. Leduc and Liu (2016) argue that a rise in uncertainty shock behaves like a negative aggregate demand shock, which drives a decline in investment expenditure. Moore (2017) finds that higher uncertainty decreases the growth in machinery and equipment investment.

Uncertainty shocks exert economic activity impacts that are of different magnitudes and lengths depending on the type of uncertainty. Extending the work of Jurado et al. (2015), Ludvigson et al. (2015) find that not all uncertainty measures are alike; a financial market uncertainty shock causes a sharp and persistent decline in the business cycle, whereas macroeconomic uncertainty responds endogenously to a business cycle shock. Supporting the Ludvigson et al. (2015) findings, Ahmadi et al. (2019) argue that oil price uncertainty has different effects on investment and that these depend on the source of uncertainty. Uncertainty in oil price driven by the global demand causes a negative response on investment. Financial market uncertainty causes a negative effect on investment with a one-year lag. Basu and Bundick (2017) argue that financial market uncertainty shocks worsen the large drop in output and investment during the Great Recession. Caldara and Iacoviello (2018) also examine the relationship between uncertainty and the U.S. economy and conclude that uncertainty in geopolitical risk causes a decline in U.S. investment for up to one year after the shock.

Meanwhile, there are only a few studies that investigate uncertainty shocks on the global oil market, particularly oil exploration. Cortazar et al. (2003) argue that uncertainty in price and geological aspect leads to the shut- down of investment in natural resources. The investment schedule is affected by current expectation of the price and by technical exposure; the schedule can be paused and resumed at any point depending on cash flow. Baumeister and Peersman (2013) contribute to the oil exploration literature by concluding that oil price volatility increases the uncertainty that leads to the postponement of investment in exploration and development. Kellogg (2014) finds and argues that firms reduce their drilling rate in response to oil price volatility.

As for the relationship between oil price and investment activity, the current research finds a positive relationship between these two measures. A rise in oil price leads to an increase in investment activity. Elder and Serletis (2010) confirm that the higher oil price induces a dramatic rise in mining expenditure. In the global oil market, Toews and Naumov (2015) apply a structural VAR and find that a 10% rise in crude oil price increases global drilling activity by 4%. Khalifa et al. (2017) apply a quantile regression framework and also argue that a positive relationship exists between oil price returns and changes in rig counts, with a lag of one quarter. Chen and Linn (2017) measure oil investment using drilling rig as the proxy, and conclude that drilling rigs respond positively to the future oil price; this however is mainly found in the regions dominated by private oil companies.

4.2.3 Contribution of the study

This study is intended to answer three research questions. First, what difference will the construction of a more refined GTU index make to knowledge about the effect of uncertainty on the global oil market? Second, what precisely do such GTU indices measure and how do they compare to the existing uncertainty indices? Third, how does oil market uncertainty affect the global oil market variables, most particularly crude oil price and oil exploration?

The GTU index is more appealing than the standard uncertainty measure used in the literature in a few ways. First, the GTU index is constructed based on the relevant terms in the oil market, which distinguishes it from the existing indices that mainly utilise macroeconomics, financial, and political variables in their construction. Second, the GTU index is constructed in a more nuanced index, accommodating supply, demand, and investment than the basic oil price index, as indicated in earlier studies. Third, the GTU data source has a more adjustable frequency (daily, weekly, monthly, and annually) and region to generate, which has more comparative advantage than the macroeconomics-based index, some of which are published quarterly. Comparing the GTU index and the financial-based uncertainty index, both can be obtained daily, but since OVX is based on oil price volatility only, it does not consider other components such as oil demand, supply, and investment. Another advantage, constructing the GTU index is less numerical and relatively simple than other indices requiring a heavy computational procedure. The GTU data source is free and publicly accessible, making it more flexible for the public to measure interest within a specific time and region. Fourth, for further empirical analysis, oil exploration is taken into account in its relationship with the GTU index and oil price to understand how uncertainty affects the oil market, as exploration responds stronger to oil price shocks than production.

The first contribution of this study is to shed light on the uncertainty literature by proposing a GTU index that is specific to the global oil market and that covers oil supply, demand, and investment. Following construction of the indices, this study compares them with the existing uncertainty indices. This research provides a more refined and specific index as a measure of oil market uncertainty measure; it thus fills the gaps in the literature on oil market uncertainty. The current approach to oil market uncertainty relies on the forecast-based index of oil price and oil inventory, or on other uncertainty indices such as macroeconomic, financial, and economic policy indices. These are not designed to directly or specifically measure how uncertainty impacts on oil market variables.

The second contribution of this study is its analysis of the performance of the GTU indices. These are constructed for basic GTU oil price and three independent components: oil supply, oil demand, and investment. The fifth and final index aggregates all components into one oil market-specific index. By accommodating these components, the GTU indices can not only encompass the critical role played by supply and demand in the oil market but can also produce a more robust estimation of what precisely the indices measure.

While the existing measures construct a Google Trends-based index for the oil market, these have significant limitations. The first index is based on a simple measure of oil price as a single component (see, for instance, Qadan and Nama (2018) and Li et al. (2019)). Their GTU indices are based on the Google Trends search term 'oil price' and its variants (e.g., 'brent'). This type of basic Google Trends index does not take into account the supply and demand factors. Therefore, the shortcomings of this approach are its restrictiveness and the possibility of bias. The second index used by the extant literature approaches the GTU oil market by references to specific events that have occurred over time. The work of Guo and Ji (2013), which is the most comprehensive work carried out so far on the GTU oil market, utilises this type of index. Their study proposes four factors as the components to construct four types of market concerns: oil price, oil demand, financial crisis, and the Libyan war. However, there are difficulties with the search terms they have chosen for the oil demand component. First, they include 'oil production' as part of the oil demand component; this is inappropriate and causes misinterpretation. Oil production is associated with the amount of oil being drilled from the oil fields for transportation, storage, and supply. Therefore, 'oil production' would be a more suitable proxy for oil supply rather than oil demand. Second, they also include 'oil consumption' in the oil demand component. Crude oil demand is not the same as oil consumption. Crude oil demand refers to the raw crude oil required for the economic activity, with much of it going to storage; oil consumption, on the other hand, encompasses the final products of refineries and other processing plants. These do not reflect raw crude oil. Hence, crude oil demand is better proxied by economic activity indicators such as industrial production, the Baltic Dry index, and Kilian (2009)'s real economic activity index. Third, the search term 'gas price' is used in the oil price component. Again, this is not suitable. It would be more appropriate to define a new component of energy price in the construction of the index rather than to include 'gas price' in the oil price category.

The choice of the Libyan war and the OPEC conference as the precautionary demand proxies by Ji and Guo (2015) also has shortcomings. Precautionary or speculative demand in crude oil is not the main motive behind the Libyan war, which was actually driven by political tension. Hence, this proxy is not directly associated with the crude oil price precautionary demand. Furthermore, OPEC conferences result in a decision to increase or cut the production quota. OPEC decisions are therefore more suitable indicators of supply shock rather than of precautionary demand shock.

The third contribution of this research is that it analyses the effect of uncertainty on the global oil market and takes into account oil exploration as a parameter of the global oil market that is extremely scarce in the existing literature. Furthermore, applying the disaggregated data for North America and world oil exploration gives more insight into how countries with dominant non-conventional oil production methods respond differently from the rest of the world to oil price shocks. This study chooses oil exploration proxied by the oil rig count to represent investment activity. The rig count is not only a measure for drilling activity, which obviously determines future oil production, it is also an important indicator widely used in the industry to measure sentiment. Oil exploration is far more responsive than oil production to a crude oil price shock; as such, it is a factor with clear importance for crude oil pricing.

The current literature focuses on the impact of global oil market shocks on uncertainty, and discusses crude oil production rather than exploration, save for one stream of literature that discusses the uncertainty effect on the macroeconomy. Kang and Ratti (2013) and Kang et al. (2017) focus on the effect of the economic policy uncertainty shocks on crude oil production. The crude oil production response is different from the oil exploration activity because it is not easy to adjust world oil production. It therefore tends to have little or no response to uncertainty shocks in the short run. It is therefore unsurprising that Kang and Ratti (2013) find that economic policy uncertainty shock does not significantly affect crude oil production. However, the U.S. oil production responds differently to world oil production, with a shortage of world production triggering the U.S. production (Kilian, 2009). Kang et al. (2017) extend Kang and Ratti (2013)'s work by disaggregating the data for U.S. and non-U.S. crude oil production. While non-U.S. oil production does not respond to an economic policy uncertainty shock, U.S. oil production responds positively in the short run to such uncertainty shocks.

Lastly, the newly proposed GTU index is an oil market uncertainty index that has worldwide scope rather being restricted to a specific country. The extant literature constructs its indices for specific countries (i.e., the U.S, Australia, Italy, New Zealand, and Turkey), which significantly reduces their generalisability, and rules them out for use in an investigation of the global measure of uncertainty.

The macroeconomic, financial, and economic policy uncertainty indices will be used as the benchmark indices in this study to compare the behaviour and performance of the newly proposed GTU index for the oil market. First, this study examines the correlation between the proposed GTU oil market indices and the benchmark indices. Second, this study investigates the response of the uncertainty benchmark indices to the global oil market shocks. Third, it examines whether the response of the GTU indices to global oil market shocks aligns with the well-established literature.

4.3 Uncertainty measures and oil market data

This study constructs the Google Trends-based Uncertainty (GTU) measure for the global oil market and then identifies how the various uncertainty measures relate to each other and to oil market data. The uncertainty measures proposed by this work are the GTU indices, and the benchmark indices are the Crude Oil Exchange Traded Fund (ETF) Volatility index (OVX), Jurado's macroeconomy uncertainty index (JMU), and the Global Economic and Policy Uncertainty (GEPU) index as the benchmark indices. All uncertainty benchmark series are depicted in Figures 4.1 and 4.2.

The OVX index is a stock market-based index of the near-term crude oil price volatility in the United States Oil Fund. The index is provided daily by the CBOE and is available from the website of the Federal Reserve Bank of St. Louis.[1] OVX measures the performance of WTI spot price changes and thus, it can be utilised as a short-term stock market uncertainty benchmark. The JMU index measures U.S. macroeconomic uncertainty based on the conditional volatility of unforeseen macroeconomy variables. This index is obtained from Jurado et al. (2015) and is updated on Ludvigson's website.[2]

The GEPU index measures the economic and policy uncertainty of 21 countries weighted by their average GDP. The chosen countries are those that contribute to two-thirds of global output. Each specific country's EPU index is measured via a process developed previously by Baker et al. (2016), and it involves counting in each country's newspaper articles the relative frequency of set terms related to 'economic', 'policy', and 'uncertainty'. The GEPU index was constructed by Davis (2016) and is updated on the website of Baker, Bloom

[1] CBOE crude oil ETF volatility indices are retrieved from FRED, Federal Reserve Bank of St. Louis; *https://fred.stlouisfed.org/series/OVXCLS*.

[2] Jurado et al. (2015) macroeconomic uncertainty indices can be retrieved from *https://www.sydneyludvigson.com/data-and-appendixes* particularly in the link: UNCERTAINTY DATA: Updated macro, real, financial uncertainty indexes 1960:07-2019:12.

and Davis.[3]

Section 4.4 illustrates the construction of the various specifications of the GTU indices. These are GTU oil price (GTU_{op}), GTU oil market-specific (GTU_{oms}), GTU oil investment (GTU_{oi}), GTU oil supply (GTU_{os}), and GTU oil demand (GTU_{od}). The raw data is obtained from Google Trends, and the GTU indices are constructed by extending Castelnuovo and Tran (2017) work specific to the oil market field.[4]

The oil market data consist of oil exploration and crude oil price. Exploration is measured by the number of oil rig counts for the world (rig_{w}) and for North America (rig_{na}), which are obtained from the Baker Hughes rig counts. [5] This research disentangles oil rig counts into North America (i.e., the U.S. and Canada), and the rest of the world. The reason behind the disaggregation is that non-conventional shale oil has dominated the U.S. and Canada since 2000, affecting both exploration and production activities. Thus, this study distinguishes the effects for countries dominated by non-conventional oil resources from those of countries with conventional oil resources.

The rig count data series for the rest of the world consists of the number of active drilling rigs for exploring or developing oil. The data cover the majority of the world's oil exploration and production, excepting the U.S. and Canada, and do not include gas and miscellaneous drilling purposes. Also excluded from the data series are specific countries characterised by civil and political tensions such as Iran, North Korea, Sudan, mainland China, Russia, the Caspian region, Sudan, Cuba, and Syria. The North America rig counts consist of U.S. and Canada oil rig counts, and again do not include gas and miscellaneous drilling.

[3]Baker, Bloom and Davis's website publish various specification of EPU indices in their website $https://www.policyuncertainty.com/index.html$. For global EPU indices, it comes from the link: Download Global EPU Data in $https://www.policyuncertainty.com/media/Global\_Policy\_Uncertainty\_Data.xlsx$.

[4]Google Trends is retrieved from $https://trends.google.com/trends/?geo = GB$.

[5]Oil rig counts are available in Baker Hughes rig count website; $https://rigcount.bakerhughes.com/static-files/7cd718fb-605d-45a0-b0f7-6a4060aff3c8$ for rest of the world and $https://rigcount.bakerhughes.com/static-files/6b00c4d1-5592-4f11-8124-3a98cb30a173$, for North America.

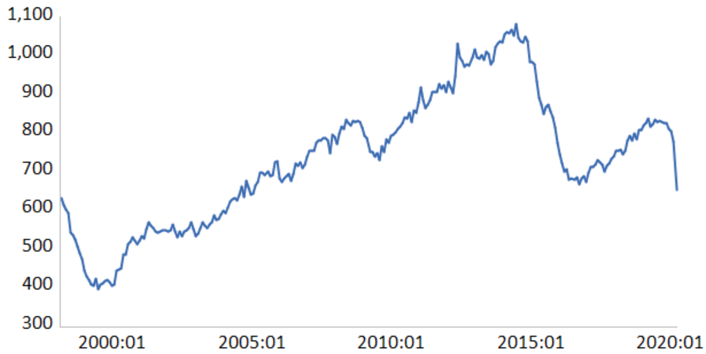
Crude oil price (op) is measured by the U.S. real acquisition cost of imported crude oil, the price being obtained from the U.S. Energy Information Administration (EIA). The series is expressed in real value after being deflated by U.S. CPI from the U.S. Bureau of Labor Statistics to incorporate the rate of inflation.[6] Figure 4.3 shows the time series of oil market variables.

The empirical analysis utilises monthly data from May 2007 to April 2021 for the OVX model and from March 1998 to April 2021 for the JMU and GEPU models. The newly proposed GTU empirical model utilises the sample period from January 2004 (when Google Trends data first became available) to April 2021. All variables are in log differences since uncertainty is associated with a short term effect. Thus, taking log differences does not cause a problem as the long-run relationship is not the main interest in this chapter.

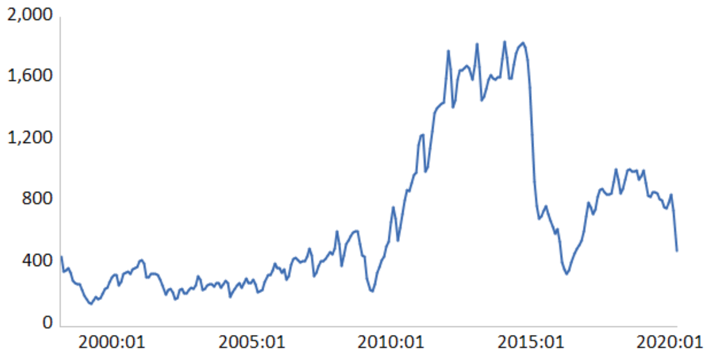
This study also applies Geopolitical Risk (GPR) and World Uncertainty Index (WUI) series for the for the purposes of testing correlation among the various uncertainty indices. The GPR index is a newspaper-based index that incorporates terms related to geopolitical conflict in 11 international papers. It is constructed by Caldara and Iacoviello (2018) at monthly intervals. The WUI index is a text mining-based index that computes the term 'uncertain' as reported by the Economist Intelligence Unit for 143 countries. The WUI index is generated by Ahir et al. (2018) at quarterly intervals. The updated GPR and WUI data series are retrieved from the website of Baker, Bloom and Davis.[7]

[6] Crude oil price is retrieved from *https : //www.eia.gov/outlooks/steo/realprices/* with the CPI data is from *https : //www.bls.gov/cpi/tables/supplemental − files/historical − cpi − u − 202005.pdf* (for all urban consumers).

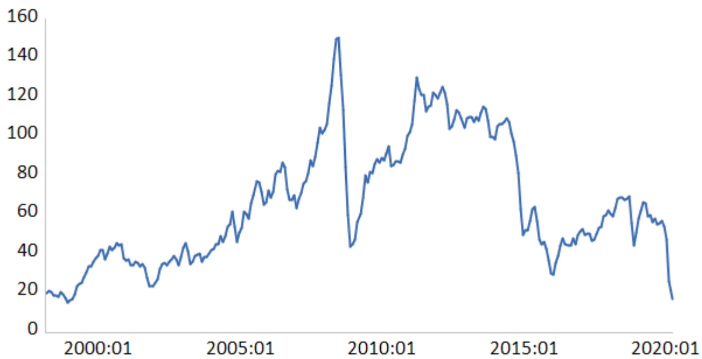
[7] GPR indices are updated in Baker, Bloom and Davis's website with the download link: *https : //www.policyuncertainty.com/gpr.html* and WUI indices are in *https : //www.policyuncertainty.com/wui$_q$uarterly.html*.



(a) World oil rig counts



(b) North America oil rig counts



(c) Crude oil prices

Figure 4.3: World oil rig counts, North America oil rig counts, and crude oil prices

4.4 Google Trends-based Uncertainty (GTU) indices: construction, comparison, and discussion

The main objective of this study is to propose GTU indices as new alternative indices for the global oil market. They measure uncertainty from freely accessible sources using a simple and relatively straightforward data treatment. The fundamental idea is when the public are interested in certain topics, particularly when there is uncertainty, they seek out specific information through a Google web search. Google Trends capture the relative frequency of specific search terms in comparison to the overall sample of search terms for a particular time and region. The higher the level of uncertainty, the higher the value of the search term in Google Trends.

The second objective of this research is to compare the trends disclosed by the newly proposed GTU indices against those of the other well-established uncertainty indices. This will be discussed in the next section. Finally, the third objective is to analyse how existing uncertainty benchmark indices and how different GTU index specifications affect the global and North American oil markets through exploration activity and crude oil price. This analysis is presented in the empirical section.

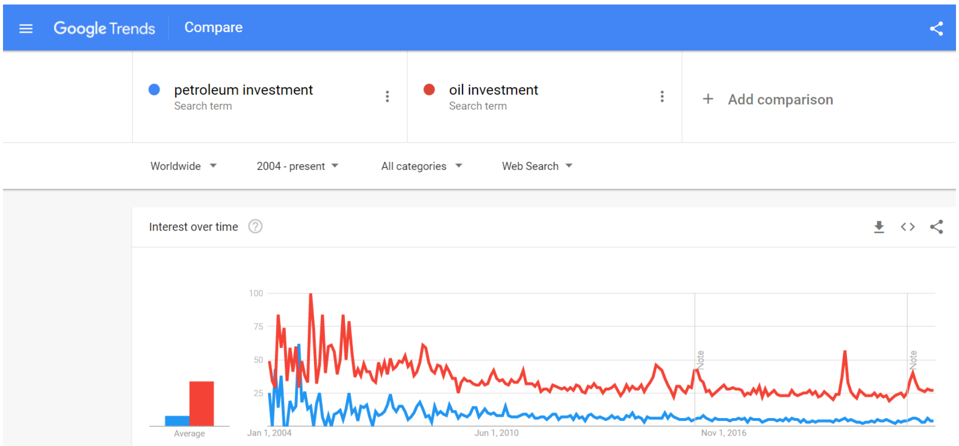
This study approaches uncertainty as the observable measure of public interest in the global oil market as demonstrated through internet access. The focus of this research is to construct a measure of uncertainty via public concern; this measure is thus more direct than the forecast-based measure. The proposed new uncertainty index is associated with current significant events that, rather than affecting future market expectation, affect the global oil market in the short-term. Google Trends shows the frequency of certain search terms relative to the total number of user queries. The more intensely the public accesses

Google to look for certain search terms indicates the level of public interest in those particular words, and thus the level of public uncertainty about them.

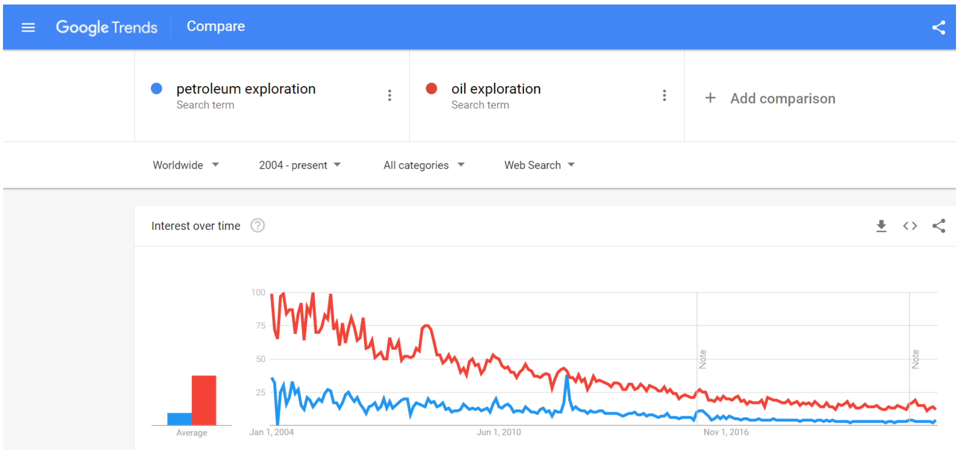
4.4.1 The newly proposed GTU indices

This section outlines the construction of the newly proposed GTU indices. There are three specifications of the GTU index proposed in this study. First, GTU oil price is proposed as the underlying index that measures public uncertainty about oil price. The GTU oil price is the most straightforward index as it applies 'oil price' as the only search term in its construction. The second index measures uncertainty for oil investment, oil supply, and oil demand. The third uncertainty measure is oil market-specific. The second and third specifications make a novel contribution to the existing literature that applies a basic index of oil price uncertainty (Guo and Ji, 2013; Qadan and Nama, 2018; Li et al., 2019). Another contribution of this research is that it utilises a basic GTU oil price as the benchmark, which it extends into more refined multi-components specific to the oil market.

There are a few reasons why the term 'oil' is preferred to 'petroleum' in obtaining Google search volume. Comparing 'petroleum investment' and 'oil investment' from January 2004 to May 2022 indicates that 'oil investment' fluctuates more than 'petroleum investment' when there are major events (i.e., a Major explosion at BP's Texas and a major earthquake hit Fukuoka, Japan in March 2005, Global Financial Crisis in June 2008, Coronavirus in April 2020) as depicted in Figure 4.4a. Secondly, comparing 'petroleum exploration' and 'oil exploration', Figure 4.4a shows the search term 'oil exploration' also fluctuated more than 'petroleum exploration', particularly in June 2008. Both search terms captured a small peak in February 2016. The search term 'petroleum exploration' has been stable, except in May 2012. Hence, in this case, using the search terms 'oil' represents more public interest by its fluctuation than using the search term 'petroleum'. Besides, petroleum is the refinery product



(a) Search term 'petroleum investment' and 'oil investment'



(b) Search term 'petroleum exploration' and 'oil exploration'

Figure 4.4: Search term 'petroleum' and 'oil' comparison in Google Trends

of crude oil, so 'oil ' is closer to the term 'crude oil' if the interest is to study crude oil exploration and investment.

The methodology of generating the second and third specifications of the indices follows the procedure by Castelnuovo and Tran (2017). Equations 4.2 and 4.3 illustrate the step, with some authorial modification of the rescaling procedure and notation. The initial step is to identify the search terms for each uncertainty specification. The choice of search terms is based on their relevance to the oil market. In order to construct GTU oil investment, GTU oil supply, and GTU oil demand, the multi-search terms described in Table 4.1 are applied, together with 'oil price' as the benchmark search term. The following specification constructs GTU oil investment, GTU oil supply, and GTU oil demand as three independent indices. The final specification aggregates these three components into a more refined and comprehensive index called GTU oil market-specific.

Table 4.1: Base Search Terms for GTU Oil investment, GTU Oil Supply, and GTU Oil Demand

Category	Search Terms			
GTU oil investment	oil investment	oil exploration	oil project	drilling
GTU oil supply	oil supply	oil production	shale oil	OPEC
GTU oil demand	industrial production	global economy	economic growth	recession

GTU index measures public interest; it is intensified when major events occur during a specific period. In uncertain times, people search on the internet more frequently, leading to the high value of the index. For example, when there is a pandemic, oil prices drop for a couple of months, and the average monthly price reached 19 USD per barrel in April 2020. The public is interested in searching for ‘oil price’, and it shows that in April 2020, Google Trends for search term ‘oil price’ is 100, which is the maximum value. Following the drop post-COVID-19, oil price increased and reached above 100 USD per barrel in March 2022, and the GTU index also peaked at 81. The other example is during Global Financial Crisis, GTU peaked in October 2008 when oil price fluctuated from 169 USD per barrel in June 2008, then a sudden drop to 49 USD per barrel in December 2008. The GTU index peaks when the oil prices fluctuate, particularly when extremely low oil prices exist. Thus, the GTU index acts as the bridge between public interest and uncertainty.

The following sub-section describes the construction of the basic GTU oil price index and the GTU indices for oil investment, oil supply, oil demand, and oil market-specific. All the indices rely on values generated by Google Trends, where $y_{i,j}$ is the relative frequency of search term i in the group of words $j, f_{i,j}$ is the highest point in a particular time and region f_j. Google Trends rescales the relative frequency value to a range of 0 to 100. The value 100 reflects the most popular term compared to other terms in the specific search group. For words that are less popular than other words, Google Trends generates their value as '< 1', while words with insufficient data are designated as 0. This study uses the 'worldwide' search region, given that it examines the global oil market. The frequency chosen is monthly frequency from January 2004 to April 2020.

$$y_{i,j} = 100 \frac{f_{i,j}}{max(f_j)} \tag{4.1}$$

4.4.1.1 GTU oil price

The basic GTU index that applies a single search term or phrase was introduced to the existing literature for the primary purpose of forecasting, particularly in macro-finance fields. Qadan and Nama (2018) examine the popularity of the single phrase 'oil price', and its variants 'crude oil' and 'price of oil' in Google Trends to reflect investor sentiment in response to the crude oil price volatility returns. Similarly, Campos et al. (2017) use the search term 'oil prices' and find that Google Trends search volume indices have a positive and statistically significant relationship with the crude oil implied volatility indices (OVX). Other research uses search terms other than 'crude oil'. For example, Li et al. (2019) apply the single term 'Brent' as the keyword and analyse its causality effect to various types of crude oil price. Afkhami et al. (2017) utilise the most popular energy commodities-related keywords from Google Trends as proxies for investor attention, and investigate their capability to predict energy commodity prices.

In this study, the basic GTU oil price will be used as the benchmark for the more advanced global oil market indices. To generate a basic index for GTU oil price, the search term i in Equation 4.1 refers to 'oil price'. In this case, it is a single search term with no comparison against other search terms. Therefore the group of words j is no longer applied. GTU oil price finds the value generated by Google Trends for 'oil price' in the 'worldwide' group

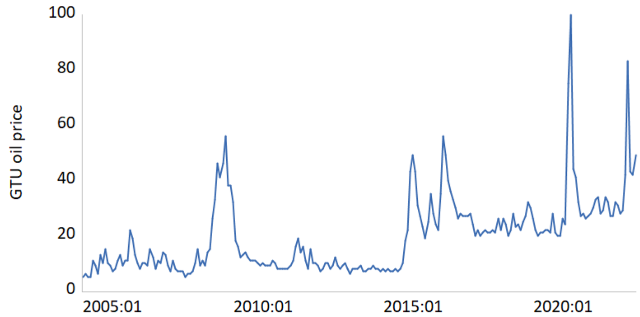
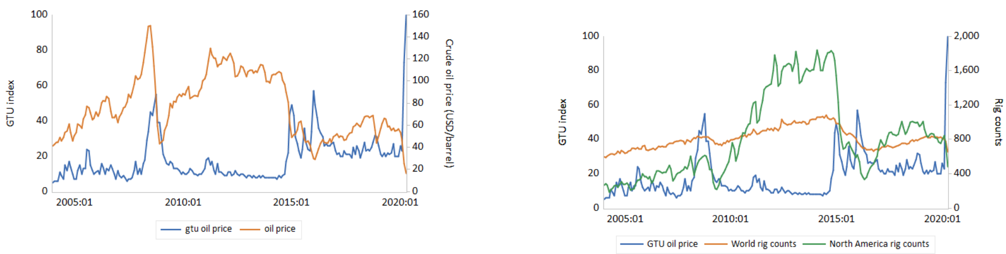


Figure 4.5: GTU oil price index



(a) GTU oil price index and crude oil price

(b) GTU oil price index and rig counts

Figure 4.6: GTU oil price index and global oil market variables

from January 2004 to April 2020 and applies this for further empirical analysis. Figure 4.5 shows that there is one peak between January 2004 to July 2014, two peaks between August 2014 to November 2019, and a final peak in April 2020. Specifically, the first peak was in October 2008 with a peak value of 55, the second and third were in January 2015 (value of 49) and January 2016 (value of 57). The last peak was also the highest, being in April 2020 with the maximum possible value of 100.

The graphical analysis, as depicted in Figure 4.6a, shows that Google Trends captures high search frequency for 'oil price' by internet users in periods when oil prices fluctuated. Crude oil price reached a high peak of 150 USD per barrel in July 2008; this was caused by strong global demand. It hit low values of 49 USD per barrel and 30 USD per barrel during the economic slowdown in January 2015 and the drop in the global stock market in January 2016 respectively, and hit the lowest point of 16 USD per barrel in April 2020, which was when the Coronavirus pandemic was reaching its peak. Hence, the GTU indices confirm that the public tend to search for more information when there is uncertainty, and this leads to the high values for the GTU oil price indices regardless of whether oil prices are high or low. The uncertainty measures in this Google Trends scope relax Killian's (2009) assumption that uncertainty is to do with speculation for the purpose of accumulating crude oil stock in readiness for a high oil price in the future, by showing that the highest level of uncertainty actually occurs when a global slump in economic activity causes the oil price to drop.

4.4.1.2 GTU oil investment, GTU oil supply, and GTU oil demand

Of the existing literature, only Guo and Ji (2013) utilise more than one component in their oil market search terms. They employ four components (oil price, oil demand, financial crisis, and Libya war) in their study, and they find a long-run equilibrium relationship between Brent oil price and the public concern for oil prices and oil demand through their Google Search Indices. One of the drawbacks of their study is that they include 'oil production' as a search term in the oil demand component. This is not appropriate given that oil production represents oil supply rather than oil demand. Furthermore, the inclusion of 'oil consumption' in the oil demand component is also not fit for purpose as oil demand does not equal oil consumption. Crude oil demand is not solely used to fuel economic activities but also counts towards inventory. Petroleum consumption, however, refers to the consumption of refined petroleum products, i.e., motor gasoline, jet fuel, kerosene, distillate fuel oil, residual fuel oil, LPG, and other petroleum liquids. Petroleum is a mixture of crude oil, natural gas, and other liquids, whereas crude oil is the liquid component of petroleum. Oil consumption refers to refined crude oil consumption ('oil' is more specific than 'petroleum'), as using petroleum will take into account the significant proportion of natural gas and other liquids in the hydrocarbon.

To construct GTU oil investment, GTU oil supply, and GTU oil demand, this study utilises the search term i from the categories provided in Table 4.1, while the group of words j refers to other search terms within that same category. Then, referring to Equation 4.2, $y_{i,j}$ is the frequency of search term i in round $j = 1$ for oil investment, $j = 2$ for oil supply, and $j = 3$ for oil demand (referencing Castelnuovo and Tran (2017), with notational adjustment by the author).

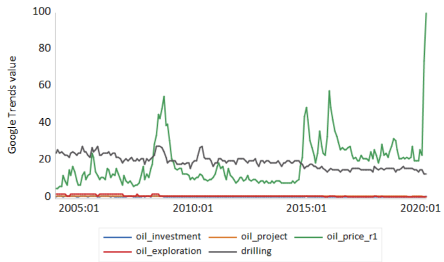
$$y_i = y_{i,j} \frac{y_b^*}{y_{b,j}} \tag{4.2}$$

This study constructs the GTU indices by applying 'oil price' as the benchmark for linking one search term with another within the same search rounds. 'Oil price' is used because it has the highest frequency and is the most volatile among these three groups. Besides, the term 'oil price' is the global oil market variable that attracts the most interest.

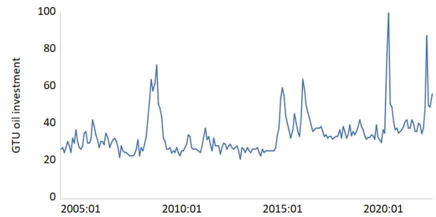
In the oil investment group, the term 'oil investment' has second-highest frequency after 'oil price', while 'OPEC' and 'global economy' are second-highest in the oil supply and oil demand groups respectively. As Google Trends only allows five search terms in each round, four search terms from the Table 4.1 categories and the benchmark term 'oil price' are input in each round. Hence, y_b^* in Equation 4.2 is the frequency of 'oil price' as a benchmark generated for GTU oil price in the previous step, $y_{b,j}^*$ is the frequency of 'oil price' in every round j, and y_i denotes the frequency of a particular search term. The monthly GTU oil market index is obtained by summing up the frequencies of a particular search term, as illustrated by Equation 4.3. The results are then rescaled to the maximum values of the GTU indices in each category in order to have the same scale for all GTU measures, as shown in Equation 4.4.

$$gtu_{j^*} = \sum_{i=1}^{N} y_i \tag{4.3}$$

$$gtu_j = 100\frac{gtu_{j*}}{max(gtu_{j*})} \tag{4.4}$$

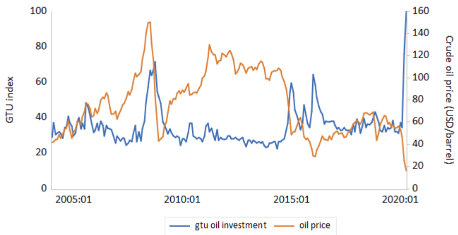


(a) The Google Trends components for GTU oil investment

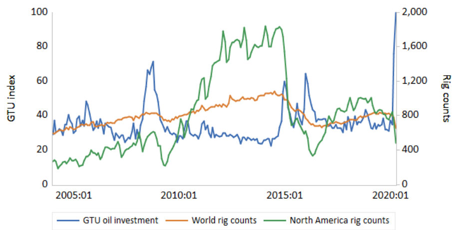


(b) GTU oil investment index

Figure 4.7: GTU oil investment



(a) GTU oil investment and crude oil price



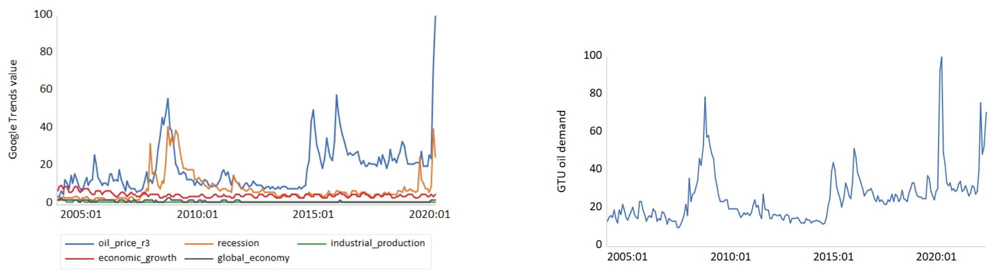
(b) GTU oil investment and rig counts

Figure 4.8: GTU oil investment and the global oil market variables

The GTU oil investment, GTU oil supply, and GTU oil demand indices exhibit fairly similar trends, showing the same major peaks over time. The highest peak is in April 2020 and there are other high peaks in October 2008, January 2015, and January 2016. The similarity derives from the dominant effect of the search term ‘oil price’. Otherwise, the terms that contribute most to these peaks are ‘oil price’, ‘recession’, ‘drilling’, and ‘OPEC’, causing the largest magnitude of GTU oil demand which is more fluctuated than GTU oil investment and oil supply. One thing that is crucial to GTU construction is to distil, from all of the available search terms, the few that best represent the category of uncertainty. It is not so much that a specific number of search terms is required, but rather that the search terms selected are popular in a certain category. As specific search terms become accessed more often, Google Trends gives their relative frequency a higher value. The interaction between a few high frequency search terms produces an uncertainty trend that is more dynamic than if many weaker search terms were used. In the case of the weak frequency terms, no matter what quantity is used, their values will not be meaningful as that of the single search term that dominates the whole sample. Hence, it is better to focus on selecting the proper search terms rather than to use a scatter-gun technique with numerous keywords that, in the end, have small or zero frequency values.

The high uncertainty in oil investment, demand, and supply almost always occur when there is a crude oil price shock, regardless of whether it is positive or negative. Although GTU indices in this study share major peaks that are similar to those of the basic GTU oil price, some of the GTU indices, depending on the uncertainty measure, exhibit peak points that are more significant and distinguishable from those of the others. Meanwhile, GTU oil price tends to have a flatter trend during periods when GTU oil investment shows a more dynamic fluctuation trend. One particular example from GTU oil investment is associated with oil exploration, such as drilling activities. Figure 4.8b explicitly shows that GTU oil investment declines when the rig counts increase. i.e., from mid–2005 to early 2008, early 2010 to mid–2014, and from 2016 until the end of 2017. On the other hand, GTU oil investment rises when the rig counts drop (from the end of 2014 to early 2016, from the end of 2017 to the end of 2018, and in April 2020). Hence, GTU oil investment can capture public interest in oil exploration when there is uncertainty. The higher the level of oil exploration, the lower the uncertainty. This aligns with Chen and Mu (2021), which provide an empirical analysis of the return-volatility relationship in 19 commodities, including crude oil. The asymmetric threshold and exponential GARCH models are applied, and their findings argue that crude oil volatility is higher following a negative demand shock, which is a leverage effect. The role of OPEC in controlling the production quota and the low marginal cost from the oil-producing countries cause the complexity of the crude oil market structure. Meanwhile, the inverse leverage effect of the positive relationship between volatility and return is found for most commodities with more inelastic demand and supply.

GTU oil demand, apart from the four significant peaks shared with basic GTU oil price, has another peak in January 2008. At this time, public uncertainty is expressed in a search for more information about the recession, which causes a spike in GTU oil demand. This peak gives the information that January 2008 is the start of the Great Recession period that followed a sharp increase in crude oil prices. Thus, GTU oil demand goes up while crude oil price drops; this may indicate that recession follows crude oil price fluctuation. Meanwhile, the GTU oil supply is less fluctuated compared to other GTU indices. GTU oil supply increases sharply in January 2015 following the drop in the world and North America rig counts.



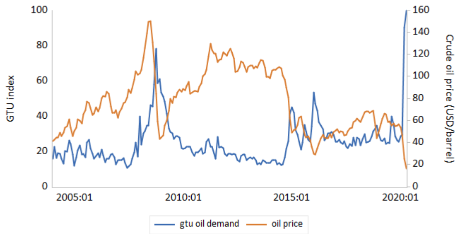
(a) The Google Trends components for GTU oil demand

(b) GTU oil demand index

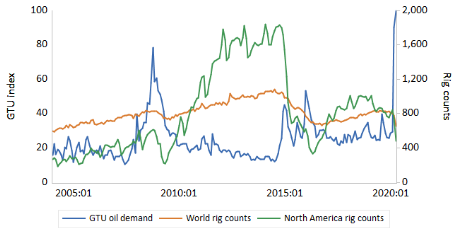
Figure 4.9: GTU oil demand

4.4.1.3 GTU oil market specific

The GTU oil market-specific index is an umbrella index that takes into account the three GTU components: GTU oil investment, GTU oil supply, and GTU oil demand, which it aggregates into a single value index. To generate

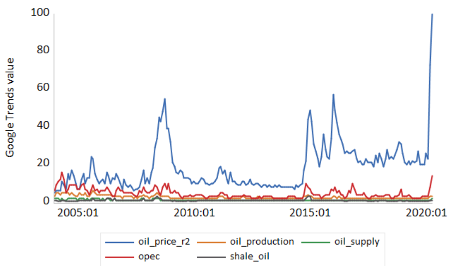


(a) GTU oil demand and crude oil price

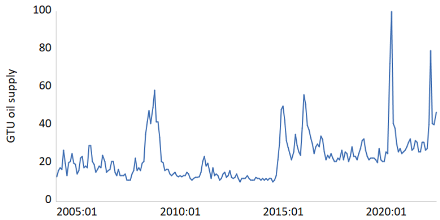


(b) GTU oil demand and rig counts

Figure 4.10: GTU oil demand and the global oil market variables

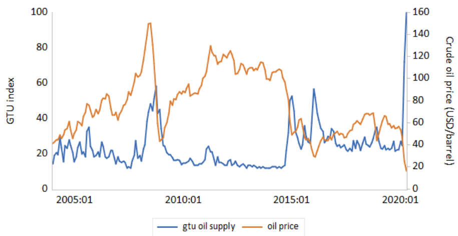


(a) The Google Trends components for GTU oil supply

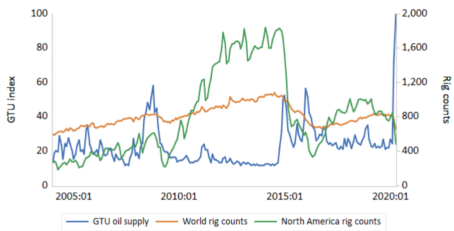


(b) GTU oil supply index

Figure 4.11: GTU oil supply



(a) GTU oil supply and crude oil price



(b) GTU oil supply and rig counts

Figure 4.12: GTU oil supply and the global oil market variables

the indices, it simply sums the values of GTU oil investment, GTU oil supply, and GTU oil demand. As shown in Equation 4.5, $j^* = 1$ refers to GTU oil investment, $j^* = 2$ is GTU oil supply, and $j^* = 3$ is GTU oil demand. Then the result is rescaled by its maximum value to the range of 0 to 100, as illustrated in Equation 4.6. Thus, all GTU indices with various specifications have the

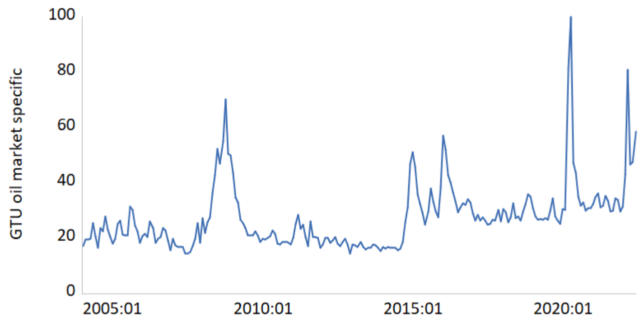
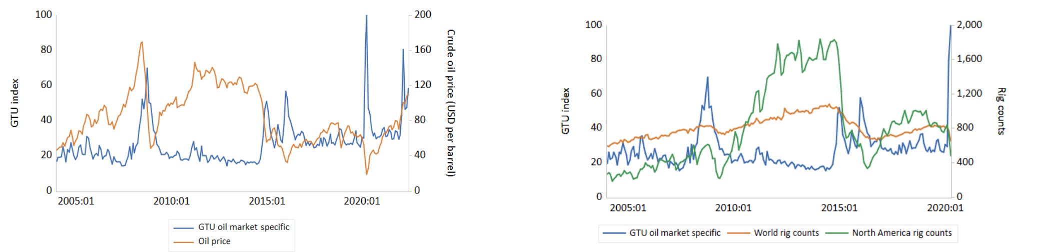


Figure 4.13: GTU oil market specific index



(a) GTU oil market specific index and crude oil price

(b) GTU oil market specific index and rig counts

Figure 4.14: GTU oil market specific index and global oil market variables

same scale.

$$gtu_{oms^*} = \sum_{j^*=1}^{N} gtu_{j^*} \tag{4.5}$$

$$gtu_{oms} = 100\frac{gtu_{oms^*}}{max(gtu_{oms^*})} \tag{4.6}$$

The three components of GTU indices interact together and form the GTU oil market-specific index. The GTU oil market-specific index shows, in addition to the main peaks already noted, fluctuation in November 2007. This fluctuation implies the beginning of a recession period, and it also shows that oil demand contributes significantly to the construction of the GTU oil market-specific index.

Overall, the most dramatic spike for all specifications of the GTU indices is in April 2020, reaching the maximum value of 100 when the Coronavirus hit

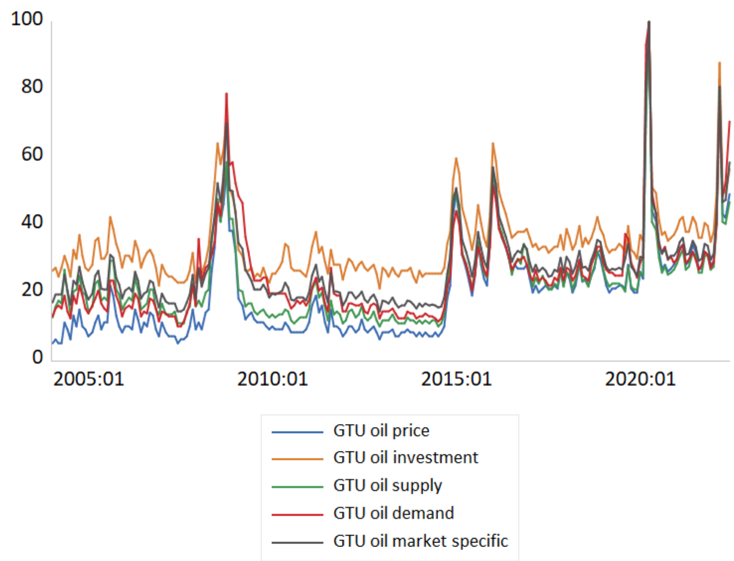


Figure 4.15: GTU indices for oil price, oil investment, oil supply, oil demand, and oil market specific

its global peak. The pandemic has exerted a tremendous impact on the world's economy. Global economic activity ceased during the various lockdowns, and companies were less likely to start a new drilling project. At this time, the crude oil price drops significantly to 16 USD per barrel, which is its lowest point in a decade, lower even than the drop during the Global Financial Crisis period. People access Google to search for 'oil price' most frequently during the Coronavirus peak that triggered the low oil price. The Coronavirus period indicates that high uncertainty is not limited to times when the crude oil price is high, but can also occur when the crude oil price is low. While the Global Financial Crisis and the Great Recession era imply that there is high uncertainty when the crude oil price peaks, it is noteworthy that the magnitude of the uncertainty is more significant during a period when the oil price is in a low state. Hence, uncertainty during this period is more reflective of the economic slowdown leading to an oil price drop rather than of speculation on future high oil price. The weak economy is associated with a drop in investment activities, and the GTU indices indicate that uncertainty rises when the rig counts drop by the end of 2014 to mid–2016, and in April 2020.

4.4.2 Descriptive statistics

The descriptive statistics of the uncertainty measures and the oil market variables in levels are shown in Table 4.2. The different specifications of the GTU indices have average values over time of between 18.37 and 35.07, with the values of GTU oil price being the lowest and GTU oil investment being the highest. There is an extensive range of minimum and maximum values that are between 5 and 100. The statistics indicate that, on average, the GTU oil investment index has up to 35% of the overall popularity of the search terms over time. The interpretation of the highest mean value of 100 indicates that in a certain period, the search term that determines the GTU index is the one that is most popular relative to the other terms in a Google web-based search. GTU oil demand and GTU oil price are slightly more volatile that the other GTU indices. Their standard deviation is higher than the other GTU specifications, while the standard deviation of GTU oil investment and GTU oil market-specific are lower than the others. When the GTU indices are compared to the benchmark uncertainty indices, the volatility of the financial market uncertainty index is moderate, in which it is comparable to the GTU indices. The geopolitical risk, global economic policy, and world uncertainty indices are more volatile than the others, while the Jurado's macroeconomy uncertainty index is the least volatile.

All the uncertainty series have long right tails and positive skewness. This is also the case for certain oil market variables in levels, i.e., world and North America rig counts, and crude oil price. The uncertainty variables have a substantial kurtosis value that indicates their distributions are more peaked than the normal distribution. They also have heavier tails than the normal distribution (i.e., they are leptokurtic) while the oil market variables have less kurtosis than a normal distribution (platykurtic) in levels. The Jarque-Bera test statistic measures the deviation of skewness and kurtosis from the normal distribution. The probability of Jarque-Bera shows the null hypothesis of normal distribution. The null hypothesis of the normal distribution is rejected at 5% significance level for all uncertainty indices and oil market variables, indicating that none of the series has a normal distribution.

Table 4.2: Descriptive Statistics (level)

	GTU_{op}	GTU_{oi}	GTU_{os}	GTU_{od}	GTU_{oms}	OVX	JMU	$GEPU$	GPR	WUI	rig	rig_{na}	op
Mean	18.367	35.072	23.061	25.208	27.601	37.380	0.968	118.951	85.552	17011.27	732.643	679.297	65.081
Median	14.500	32.900	21.250	22.556	25.398	33.150	0.953	105.899	66.164	14223.15	740.500	461.875	58.022
Max	100.000	100.000	100.000	100.000	100.000	154.437	1.323	344.158	545.094	59100.97	1080.000	1840.750	150.114
Min	5.000	22.349	11.667	10.526	15.316	16.678	0.882	47.224	23.702	4396.794	395.000	137.600	14.697
Std.dev	12.574	10.301	11.578	12.583	11.180	17.429	0.065	55.740	63.745	8802.704	169.574	485.020	31.188
Skewness	2.500	2.601	2.579	2.668	2.721	3.290	2.544	1.372	3.012	1.482	0.007	1.028	0.517
Kurtosis	13.102	12.759	13.888	13.386	14.247	19.289	11.329	5.045	16.548	5.933	2.304	2.830	2.288
Jarque-Bera	1037.563***	998.713***	1185.396***	1113.414***	1274.895***	2006.011***	1055.819***	129.756***	3884.151***	261.594***	5.376**	47.204***	17.455***

4.4.3 Correlation between GTU indices and other uncertainty benchmark measures

This section examines how the different measures of uncertainty are correlated with the GTU indices, the objective being to find the benchmark for modelling the relationship between uncertainty and the crude oil price and exploration activity variables in the next section. The correlation between the newly proposed GTU indices and the various other uncertainty benchmarks (i.e., financial market, macroeconomy, global economic and policy, geopolitical risk, and world uncertainty) are investigated as shown in Table 4.3. The GTU indices have a positive correlation with the other uncertainty measures, and the correlation ranges are from 0.02 to 0.83. Overall, the various specifications of the GTU indices are highly correlated with the financial market and macroeconomy indices, and less correlated with the economic and policy, geopolitical risk, and world uncertainty indices.

The GTU indices are positively and highly correlated with the oil volatility index, with a range of 0.63 to 0.83. The time series plot of GTU and OVX also shows peaks points that are nearly the same over time. The correlation ranges between GTU and the JMU macroeconomy uncertainty indices are between 0.42 and 0.74. The positive correlation between the GTU indices and these two mainstream uncertainty indices is consistent with the existing literature, as is their magnitude. Jurado et al. (2015) examine the correlation between their macroeconomy index and the oil volatility index, and find a positive correlation with the value of about 0.50. An interesting finding is that Castelnuovo and Tran (2017) construct a GTU index for macroeconomic uncertainty and find that it too has a high correlation (0.63) with Jurado's macroeconomy indices. This implies that the different specifications of the GTU indices have a consistently higher and positive correlation with Jurado's macroeconomy indices. The GTU indices also have a positive correlation with

GEPU, with ranges of 0.12 to 0.27. The GTU indices have a positive and low correlation with the geopolitical index, with ranges of 0.05 to 0.26. Meanwhile, the world uncertainty index is negatively correlated with the GTU, oil volatility, and macroeconomy uncertainty indices.

Based on graphical analysis, Figure 4.16 depicts the time series plot between the GTU indices and the other uncertainty measures. In this case, the GTU oil market-specific index is taken to illustrate the correlation with the benchmark uncertainty indices. Across the measures, the GTU oil market-specific, financial market, macroeconomy, and global economic policy indices peak in the Global Financial Crisis (GFC) period whereas the geopolitical risk and world uncertainty indices do not. This indicates that the GTU, macroeconomy, economic policy, and financial market indices are sensitive to the global economy and can capture the economic uncertainty that occurred between 2008 and 2009 very well. Aside from the GFC period, the GTU indices mimic the financial market uncertainty index until the next peak point in 2015 and 2016. This point occurs during a period of sluggish global demand and low crude oil price. The last outbreak point for both indices occurs in March to April 2020, which correlates with the Coronavirus pandemic.

The time series plot also indicates that even if the indices have a similar movement trend over time, the GTU indices move before the benchmark financial uncertainty index, save for a few lags. Both indices communicate that public interest is high during the peak point, triggering the public to gather information as quickly as possible. But Google Trends can be accessed freely at any time, which means that its trend movement is quicker compared to that of the financial market uncertainty index, which can only be derived after particular trading occurs. A similar behaviour pattern is found in the GTU and Jurado's macroeconomy time series plot. They both experience peaks in 2008, 2015, 2016, and 2020, but the GTU series moves a few lags before the macroeconomy uncertainty index. This is because the macroeconomy index involves many macroeconomic variables where the values are generated in response to (and therefore after) the breakpoint.

The time series trend then also confirms the positive and high correlation between GTU and the two other benchmark indices, the financial market and macroeconomy uncertainty indices, as shown in Table 4.3. These measures approach uncertainty with different measurement parameters and objectives. However, they are interconnected in that they aim to explain the uncertainty evinced towards major dramatic events that affect economic and financial circumstances. Therefore, it is expected that there is a positive correlation and similar trend between these uncertainty indices, particularly in the wake of unprecedented major events that lead to an economic slowdown.

Neither the geopolitical risk nor the world uncertainty indices are sensitive to the GFC period. Initially, the geopolitical risk index is fairly stable, then it exhibits an upward trend line from 2013 to 2018; this corresponds to the Syrian war and the Paris attacks. The world uncertainty indices show a few spikes in 2012, 2015, and 2017 which correspond to the European crisis, Brexit, and U.S. presidential election, respectively. In support of the low and sometimes negative correlation results, the plots indicate that there are no co-movements between the GTU indices and either the geopolitical risk or the world uncertainty indices. One possible reason for this is that these two indices are negatively correlated with the other benchmark indices, probably because they measure the general term of 'uncertainty' or they are focused on specific events that have a broad scope and exert little to no effect on the economy generally, and more specifically on the oil market.

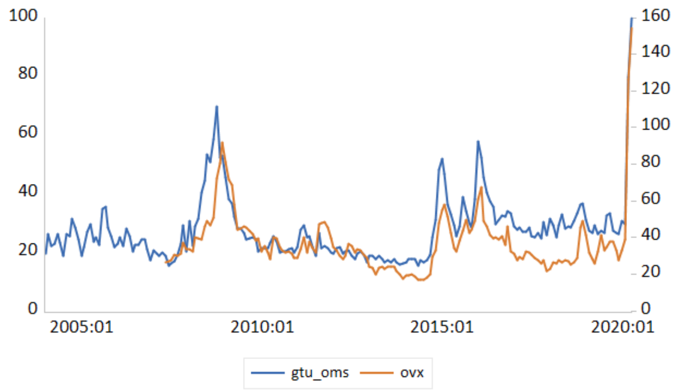
The empirical analysis will therefore focus on the financial market (OVX), Jurado's macroeconomy (JMU), and the global economic policy (GEPU) indices as the benchmark models and will analyse whether the newly proposed GTU indices behave consistently with these benchmark indices. The more consistent the behaviour of the GTU indices regarding the oil market variables, the more confident the literature can be in utilising the GTU indices as an alternative new uncertainty measure.

Table 4.3 illustrates that the GTU and the benchmark uncertainty measures have a negative correlation with oil rig counts. The magnitudes of the correlation for GTU indices to world oil rig counts are from 0.28 to 0.39. The GTU indices have a slightly higher correlation with the North America rig counts (i.e., from 0.37 to 0.50). These ranges are in line with the financial market and macroeconomy uncertainty correlations with oil rig counts, which are between 0.28 and 0.58. The correlations between all uncertainty measures and the crude oil price are negative. The range of the correlation between the various specifications of the GTU indices and oil price is between 0.32 and 0.53, which is comparable to the correlation of the financial market and global economic policy uncertainty indices with the oil price.

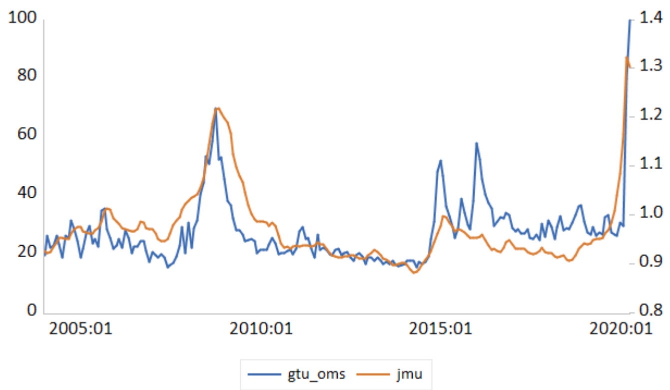
The rig counts have a positive correlation with the crude oil price, being 0.56 for world rig counts and 0.48 for the North America rig counts. Various specifications of the GTU indices have a high correlation to each other, ranging from 0.85 to 0.99. The high correlation among GTU indices is caused by the dominant effect of the search term 'oil price', which plays an essential role as the benchmark even if other search terms are less popular than the benchmark. By way of illustration, 'recession' is applied as one of the search terms in GTU oil demand and shows higher popularity among other terms within the same group. Thus, the search term 'recession' is dominant to other search terms in the oil demand component. But when the correlation between GTU oil price and GTU oil demand is examined, it shows that GTU oil price is 85% correlated with GTU oil demand which is less than the correlation between GTU oil price and the other GTU indices. This indicates that since both 'oil price' and 'recession' search terms are dominant, the oil price correlation is not as strong as when 'oil price' is linked with another less-dominant search term. Then, it can be concluded that two dominant search terms interacting with each other cause lower correlation than when a single dominant search term is utilised.

Table 4.3: Pairwise correlation between uncertainty indices and oil market variables (in level)

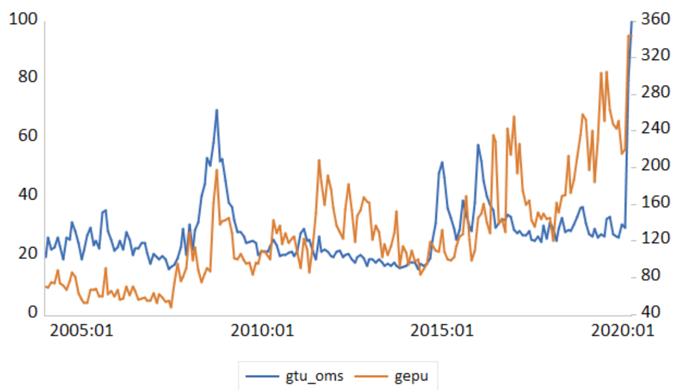
	GTU_{op}	GTU_{oi}	GTU_{os}	GTU_{od}	GTU_{oms}	OVX	JMU	$GEPU$	GPR	WUI	rig	rig_{na}	op
GTU_{op}	1												
GTU_{oi}	0.955	1											
GTU_{os}	0.985	0.966	1										
GTU_{od}	0.854	0.870	0.862	1									
GTU_{oms}	0.951	0.974	0.972	0.951	1								
OVX	0.631	0.676	0.655	0.833	0.756	1							
JMU	0.424	0.541	0.469	0.743	0.617	0.748	1						
$GEPU$	0.269	0.123	0.192	0.211	0.186	0.082	-0.102	1					
GPR	0.258	0.102	0.189	0.054	0.116	-0.173	-0.176	0.488	1				
WUI	0.082	-0.041	0.025	-0.030	-0.016	-0.082	-0.112	0.620	0.375	1			
rig	-0.365	-0.277	-0.347	-0.385	-0.354	-0.278	-0.286	-0.237	-0.195	-0.048	1		
rig_{na}	-0.389	-0.374	-0.399	-0.500	-0.447	-0.466	-0.576	0.002	-0.029	0.040	0.855	1	
op	-0.531	-0.317	-0.460	-0.416	-0.417	-0.325	-0.089	-0.412	-0.464	-0.316	0.557	0.477	1



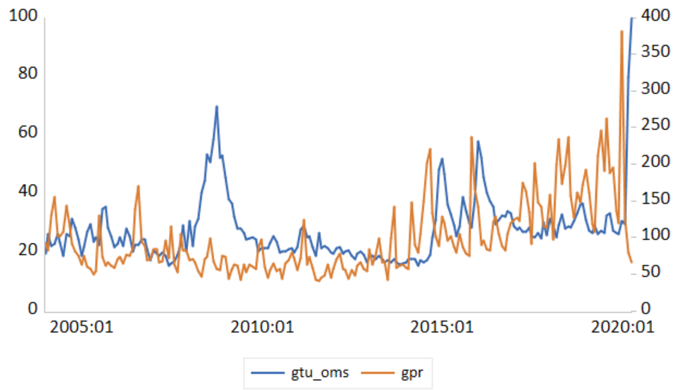
(a) GTU Oil Market Specific and Oil Volatility Indices



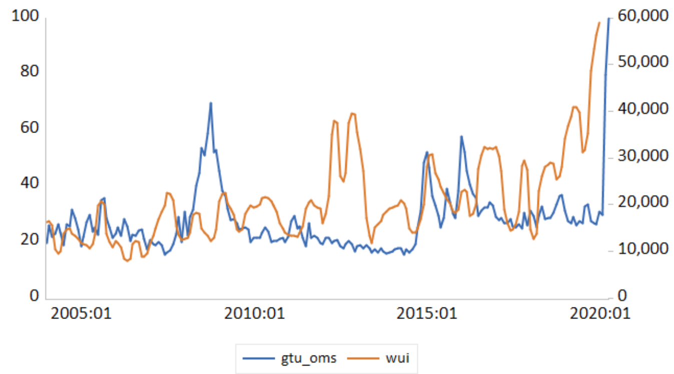
(b) GTU Oil Market Specific and Jurado's Macroeconomy Indices



(c) GTU Oil Market Specific and Global Economic Policy Indices



(d) GTU Oil Market Specific and Geopolitical Risk Indices



(e) GTU Oil Market Specific and World Uncertainty Indices

Figure 4.16: The GTU Oil Market Specific Indices and Other Existing Uncertainty Measures

4.5 Methodology

This study applies a VAR framework to investigate the effects of uncertainty on two global oil market variables: oil exploration and crude oil price. Augmented Dickey-Fuller (ADF) and Enders and Lee (EL) Fourier unit root tests are employed, using the detailed methodology described in the previous chapter on crude oil price and oil production (Chapter 3, Subsection 3.3.3). Enders and Lee (2012) unit root test has distinct features compared to other tests, i.e. Zivot and Andrews (1992) and Narayan and Popp (2010), such as the flexibility with the break form to accommodate the either smooth or sharp break, no need to know how many breakpoints, and no need to know when the breaks occur. Zivot and Andrews (1992) is too restrictive that it can only deal with one break point in the series. Narayan and Popp (2010) has the flexibility as Enders and Lee (2012) test in terms of no prior knowledge when the breaks occur. However, Narayan and Popp (2010) test limits the number of the breaks into two breaks in level and slope. For comparison purposes, the Narayan and Popp (2010) unit root test is presented in Appendix B.2.

The stationarity conclusion drawn based on ADF and EL tests follows these procedures: determining the optimal frequency and lag length, using the F-test to test the null hypothesis of linearity, and using the EL test result to conclude the unit root properties of the data series if the F-test statistic rejects the null hypothesis; otherwise, the standard ADF test is used to conclude. The critical values of F(k) single frequency refers to Enders and Lee (2012), Table 1, Panel c.

4.5.1 VAR framework

VAR estimation is conducted to estimate the effects of uncertainty on oil exploration and crude oil price. The causality effect is estimated through the Granger Causality test. Then, impulse response and variance decomposition are generated to predict the contribution of uncertainty shocks to the variability of oil exploration and crude oil price. The VAR model takes into account the past values of the dependent variable and the other variables to estimate the current value of the variable of interest, and it treats all variables as endogenous. The reduced form VAR model with a lag from p=1 up to lag q can be represented by the equation:

$$x_t = A_0 + \sum_{p=1}^{q} A_1 x_{t-p} + e_t \tag{4.7}$$

In a matrix algebra representation, equation 4.7 can be written as follows.

$$\begin{bmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{bmatrix} = \begin{bmatrix} A_{10} \\ A_{20} \\ A_{30} \end{bmatrix} + \begin{bmatrix} A_{11}(L) & A_{12}(L) & A_{13}(L) \\ A_{21}(L) & A_{22}(L) & A_{23}(L) \\ A_{31}(L) & A_{32}(L) & A_{33}(L) \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \\ e_{3t} \end{bmatrix} \tag{4.8}$$

where x_t is a 3 x 1 vector of endogenous variables: $x_t = \begin{bmatrix} x_{1t}, x_{2t}, x_{3t} \end{bmatrix} = \begin{bmatrix} rig_t, op_t, unc_t \end{bmatrix}'$. A_{i0} is the element i of vector A_0 which is a 3 x 1 matrix of intercept terms. $A_{ij}(L)$ are the elements in row i and column j of the matrix $A - 1$, which are 3 x 3 matrices of lag coefficients in the lag operator L. e_{it} is the element i of the vector e_t, a 3 x 1 matrix of white noise disturbances. The variable rig_t refers to the time series of world rig count rig_w and North America rig count rig_{na} and the unc_t refers to each of uncertainty indicators; ovx, jmu, gepu, gtu_{op}, gtu_{oi}, gtu_{os}, gtu_{od}, and gtu_{oms}. Each equation has zero expected value of error: $E(e_{it}) = 0$.

4.5.1.1 Granger Causality test

The term $A_{ij}(L)$ indicates the coefficients of variable j and the lagged values of variable i which can be expressed into individual elements $a_{ij}(1), a_{ij}(2)$, $a_{ij}(3), ..., a_{ij}(q)$ for lag $p = 1$ up to $p = q$. Variable j does not Granger cause variable i if all coefficients of the $A_{ij}(L)$ are equal to zero. In this setting, $a_{i_1j_1} = a_{rig\,op}, a_{i_1j_2} = a_{rig\,unc}, a_{i_2j_1} = a_{op\,rig}, a_{i_2j_3} = a_{op\,unc}, a_{i_3j_1} = a_{unc\,rig}$, and $a_{i_3j_2} = a_{unc\,op}$, where rig is rig counts (world and North America), op is crude oil price, and unc is the uncertainty index (benchmark and GTU indices).

$$a_{ij}(1) = a_{ij}(2) = a_{ij}(3) = ... = a_{ij}(q) = 0 \tag{4.9}$$

Hence, the null hypothesis is $H_0 : A_{ij}(L) = 0$. The Granger causality direction is from variable j to variable i. The null hypothesis can be rejected if there are one or more values in Equation 4.9 that are not zero significantly. The rejection of null hypothesis means that the past value of variable j helps to forecast the current value of variable i. This chapter also considers the structural break and formally tests the break applying Supremum Wald (swald) and Supremum Likelihood-Ratio (slr) as described in Subsection 2.4.2.2. If the short-run parameter stability of the null hypothesis is rejected, then the time-varying Granger causality test is applied.

The VAR-based time-varying Granger causality are sequences of Wald statistics and bootstrapped-based critical values as proposed by Shi et al. (2018, 2019). The time-varying Granger causality test, Wald statistics is calculated from the subsamples of the observation. Referring to the VAR(q) expression in Equation 4.7, the Wald statistic imposed by the null of no Granger causality from variable x_1 to x_2 is defined as

$$W = [R\,vec(\hat{A})]'[R(\hat{\Omega} \otimes (X'X)^{-1}R')]^{-1}[R\,vec(\hat{A})], \tag{4.10}$$

where R is the coefficient restriction matrix and $vec(\hat{A})$ is the row vectorised coefficients of $\hat{A}$. Let the total number of observations as T, the starting point of the regression sample be f_1, the ending point is f_2, and $f_w = f_2 - f_1$. The Wald statistic is denoted by $W_{f_1}^{f_2}$, where $\tau_1 = [f_1 T], \tau_2 = [f_2 T], \tau_w = [f_w T]$, and the minimum required observations to estimate VAR is $\tau_0 = [f_0 T]$. There are three procedures to detect non-constant causal relationship and generate the sequences of statistics: forward expanding, rolling, and recursive evolving windows. This study applies rolling window procedure, suggested by Swanson (1998) which its start point is $\tau_1 = \tau_2 - \tau_0 + 1$ and the end point $\tau_2 = \tau_1 + \tau_0 - 1$. The window size is assumed to be fixed and it equals to τ_0. The test statistic is the supremum Wald statistic sequence and is expressed as

$$SW_f(f_0) = \sup_{f_2 = f, f_1 \in [0, f_2 - f_0]} (W_{f_1, f_2}). \tag{4.11}$$

4.5.1.2 Impulse response function

Impulse response function is used to examine the effects of a one time shock on all the current and future endogenous variables through the dynamic of the VAR. Impulse response can be obtained by transforming the VAR as Equation 4.8 into the Vector Moving Average (VMA) representation (Enders, 2015).

$$\begin{bmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{bmatrix} = \begin{bmatrix} \bar{x_1} \\ \bar{x_2} \\ \bar{x_3} \end{bmatrix} + \sum_{n=0}^{\infty} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}^i \begin{bmatrix} e_{1t-i} \\ e_{2t-i} \\ e_{3t-i} \end{bmatrix} \tag{4.12}$$

The vector of errors e_t can be written in the form of the sequences of the shocks ε_t, where $e_t = B^{-1}\varepsilon_t$,

$$x_t = \bar{x} + \sum_{n=0}^{\infty} A_1^n B^{-1}\varepsilon_t \tag{4.13}$$

$$\begin{bmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{bmatrix} = \begin{bmatrix} \bar{x_1} \\ \bar{x_2} \\ \bar{x_3} \end{bmatrix} + \sum_{n=0}^{\infty} \begin{bmatrix} \phi_{11}(n) & \phi_{12}(n) & \phi_{13}(n) \\ \phi_{21}(n) & \phi_{22}(n) & \phi_{23}(n) \\ \phi_{31}(n) & \phi_{32}(n) & \phi_{33}(n) \end{bmatrix} \begin{bmatrix} \varepsilon_{1t-n} \\ \varepsilon_{2t-n} \\ \varepsilon_{3t-n} \end{bmatrix} \tag{4.14}$$

and where $\phi_{ij}(n) = A_1^n B^{-1}$,

$$x_t = \bar{x} + \sum_{n=0}^{\infty} \phi_n \varepsilon_{t-n} \tag{4.15}$$

The coefficients of $\phi_{ij}(n)$ are the impulse response functions, and a visual interpretation can be obtained by plotting the $\phi_{ij}(n)$ against i.

As the information in matrix A_1 is not sufficient for identifying all parameters in the VAR system, the model requires the imposition of additional restrictions to identify the impulse responses. One of the identification restrictions is the Choleski decomposition, which decomposes the residuals in the triangular form. A theoretical assumption is required to determine which variable has no contemporaneous effect on other variables and also the consequent order of the variables. Furthermore, it is essential to understand that the ordering procedure is based on the correlation coefficients between e_t. Thus, if the correlation coefficient between e_{1t} and e_{2t} is ρ_{12}, the significance of ρ_{12} may be tested by specifying the null hypothesis of no correlation between e_{1t} and e_{2t}; $h_0 : \rho_{12} = 0$. If ρ_{12} is statistically significant, the impulse response function is obtained through that particular ordering (Enders, 2015). To test the robustness of the model, this study also does reordering of the variables, with the results being available in Appendix B.5.

4.5.1.3 Variance decomposition

Variance decomposition describes the variation of each endogenous variable into the shock components of other variables in the VAR system. It provides information as to the proportion to which the shocks contribute to one endogenous variable, including the variable's own shocks as well as those of the other variables. The n-step-ahead forecast error is the difference between the realisation and the forecast value.

$$x_{t+n} = \mu + \sum_{i=0}^{\infty} \phi_i \varepsilon_{t+n-i} \tag{4.16}$$

$$x_{t+n} - E_t x_{t+n} = \sum_{i=0}^{n-1} \phi_i \varepsilon_{t+n-i} \tag{4.17}$$

For the x_{1t} sequence:

$$\begin{aligned} x_{1t+n} - E_t x_{1t+n} = {} & \phi_{11}(0)\varepsilon_{x_1 t+n} + \phi_{11}(1)\varepsilon_{x_1 t+n-1} + ... + \phi_{11}(n-1)\varepsilon_{x_1 t+1} \\ & + \phi_{12}(0)\varepsilon_{x_2 t+n} + \phi_{12}(1)\varepsilon_{x_2 t+n-1} + ... + \phi_{12}(n-1)\varepsilon_{x_2 t+1} \\ & + \phi_{13}(0)\varepsilon_{x_3 t+n} + \phi_{13}(1)\varepsilon_{x_3 t+n-1} + ... + \phi_{13}(n-1)\varepsilon_{x_3 t+1} \end{aligned} \tag{4.18}$$

Hence, the n-step-ahead forecast error of x_{1t+n} is $\sigma_{x1}(n)^2$, which consists of the shocks of ε_{x_1t}, ε_{x_2t}, and ε_{x_3t} as follows.

$$\begin{aligned}\sigma_{x1}(n)^2 = \sigma_{x1}(n)^2 &\left[\phi_{11}(0)^2 + \phi_{11}(1)^2 + ... + \phi_{11}(n-1)^2\right] \\ &+ \sigma_{x2}(n)^2 \left[\phi_{12}(0)^2 + \phi_{12}(1)^2 + ... + \phi_{12}(n-1)^2\right] \\ &+ \sigma_{x3}(n)^2 \left[\phi_{13}(0)^2 + \phi_{13}(1)^2 + ... + \phi_{13}(n-1)^2\right] \qquad (4.19)\end{aligned}$$

4.6 Empirical Results and Analysis

This study applies reduced form VAR to analyse the effects of uncertainty on oil exploration activity and the crude oil price. The main section of the empirical results deals with Granger causality, impulse response, and variance decomposition of the oil market to uncertainty shock. Prior to undertaking VAR analysis, a unit root test is conducted. The optimal lag length selection for each model (ranging from three to twelve lags) is described in Appendix B.1 and the significance of the variables in the VAR equations is reported in Appendix B.3. An autocorrelation test is performed to ensure the model is free from serial correlation at the chosen lag, and the results are shown in Appendix B.4. The order of the variables in the base model of VAR is as follows: rig counts, oil price, and uncertainty index. Oil rig count does not respond contemporaneously to crude oil price and uncertainty shocks because there is a lag in the time required to adjust the exploration project. Oil price and uncertainty do however receive contemporaneous effects from exploration activity changes, and uncertainty responds contemporaneously to oil price and rig count changes. For robustness test purposes, Brent price is applied as the proxy for the crude oil price (see Appendix B.5), and the variables in the VAR estimation are reordered into rig counts, uncertainty, and oil price.

4.6.1 Unit Root Test

A unit root test is performed to examine the stationarity of each series before estimating the relationship between the uncertainty measures, both benchmark and newly proposed GTU, and the global oil market variables. Table 4.4 shows that the ADF unit root test using intercept only results in non-stationarity at 5% significance level for all variables in levels, except for the JMU index. Meanwhile, the ADF test with intercept and trend suggests stationarity for the GEPU and JMU indices. The results for JMU are as expected, given that Jurado et al. (2015) construct their macroeconomy uncertainty indices to be stationary in levels.

In addition, the EL test is carried out to control for structural breaks in the crude oil price, rig count, and uncertainty index time series. Enders and Lee (2012) unit root test have distinct features compared to other tests (i.e. Zivot and Andrews (1992); Narayan and Popp (2010)), such as the flexibility with the break form to accommodate either smooth or sharp break, no need to know how many breakpoints, and no need to know when the breaks occur. Zivot and Andrews (1992) are too restrictive that the test can only deal with one break point in the series. Narayan and Popp (2010) have the flexibility as Enders and Lee (2012) test in terms of no prior knowledge when the breaks occur. However, Narayan and Popp (2010) test limit the number of the breaks into two breaks in level and slope. Narayan and Popp (2010) indicate stationary results for all variables at level, including oil price, which is a sceptical result, as oil prices have been through many fluctuation periods. The Narayan and Popp (2010) test result is presented in Table B.2, Appendix B.2.

The ADF and Enders and Lee (2012) result is presented in Table 4.4. Comparing the F-stat and the critical values of $F(k)$ single frequency in Table 1, Panel c, Enders and Lee (2012), F-stat with a null hypothesis of linearity is rejected for OVX, world rig counts, and oil price. Thus, for these variables, Enders and Lee (2012) test in Panel (b), Table 4.4, is used to examine the unit root properties of the data series, while the rest of the variables apply the standard ADF test provided in Panel (a), Table 4.4. The null hypothesis of the Enders and Lee (2012) test is that there is a unit root with unknown breakpoints. Following these steps, the final result generates the mixed stationary of variables at level, with OVX, North America rig counts, and oil price being non-stationary. The results of both tests suggest that all variables are stationary in first difference and log first difference.

Based on Figure 4.3, the world oil rig counts exhibit structural breaks

Table 4.4: Unit root test for uncertainty and oil market variables

(a) ADF t-statistics

	Level	Log-level	First difference	Log-first difference
JMU	-3.601**	-3.506**	-10.748***	-10.470***
OVX	-3.477***	-4.195***	-9.917***	-8.987***
GEPU	-3.407**	-3.670**	-6.411***	-8.554***
GTU_{op}	-4.478***	-4.033***	-9.652***	-15.997***
GTU_{oi}	-5.738***	-4.820***	-9.589***	-12.317***
GTU_{od}	-3.681**	-3.763**	-9.695***	-9.534***
GTU_{os}	-4.563***	-4.156***	-9.864***	-12.167***
GTU_{oms}	-4.279***	-4.122***	-9.809***	-12.112***
rig_{w}	-2.090	-2.399	-4.992***	-4.789***
rig_{na}	-2.620	-2.329	-4.812***	-5.141***
op	-3.470**	-2.584	-10.808***	-15.849***

(b) Enders and Lee (2012) t-statistics

	Level	Log-level	First difference	Log-first difference
JMU	-4.268**	-4.285**	-10.336***	-6.267***
OVX	-2.657	-1.886	-6.594***	-4.236***
GEPU	-5.386***	-4.207***	-10.352***	-8.436***
GTU_{op}	-4.650***	-3.803***	-9.112***	-6.969***
GTU_{oi}	-4.522***	-4.745***	-8.946***	-8.803***
GTU_{od}	-2.513	-2.325	-4.334**	-9.305***
GTU_{os}	-4.681***	-3.934***	-9.731***	-6.009***
GTU_{oms}	-3.226**	-3.917***	-12.770***	-9.256***
rig_{w}	-4.211**	-4.513**	-5.050***	-4.788***
rig_{na}	-2.635	-3.951*	-7.560***	-7.546***
op	-3.120	-3.370*	-6.413***	-5.104***

(c) Enders and Lee (2012) F-statistics

	Level	Log-level	First difference	Log-first difference
JMU	3.831	3.785	3.007	3.036
OVX	4.071**	3.262***	14.238***	0.662
GEPU	5.168	3.715***	11.940***	0.756
GTU_{op}	2.765	2.361	4.115	1.702
GTU_{oi}	2.658	2.133	3.636	2.155
GTU_{od}	5.078	3.074***	14.319***	2.358
GTU_{os}	2.827	2.440	1.285	1.313
GTU_{oms}	3.949	2.670***	6.256	1.591
rig_{w}	8.396***	4.196	7.048***	0.809
rig_{na}	6.550	9.327**	1.122	0.533
op	4.084***	3.247	5.994	4.885

in trend while the crude oil price series indicates breakpoints in the intercept. These possible breaks may cause bias when using the ADF test to confirm the stationarity results. The Enders and Lee (2012) test, being able to overcome changes in either intercept or trend due to possible breaks over time, would be more appropriate. This study treats all uncertainty indices, the rig counts, and the crude oil price as stationary in log first difference to analyse the short-term relationship among the variables. Kilian and Murphy (2014) define the characteristic of a market concern shock (i.e., one related to speculative demand) is that it has a large and immediate effect on oil price and a temporary effect on economic activity.

Table 4.5: Structural break test result

Model	Equation	World rig counts		North America rig counts	
		swald	slr	swald	slr
JMU	rig	42.18**	44.62**	105.18***	111.81***
	op	74.96***	75.44***	110.22***	116.46***
	JMU	41.03**	43.49**	97.17***	104.33***
OVX	rig	76.06***	77.28***	92.77***	115.15***
	op	51.73***	56.06***	104.75***	125.60***
	OVX	27.51	32.03	88.61***	111.36***
GEPU	rig	34.68	37.30	126.61***	133.68***
	op	55.20***	57.36***	111.33***	120.42***
	GEPU	25.29	27.66	44.96	54.75
GTU_{op}	rig	68.65***	70.47***	83.40***	100.54***
	op	54.96***	58.14***	107.66***	122.56***
	GTU_{op}	41.24**	45.03**	68.59**	85.87***
GTU_{oi}	rig	71.85***	75.44***	83.49***	100.63***
	op	47.42**	52.66***	73.37***	90.72***
	GTU_{oi}	35.07	40.14	58.50	75.24***
GTU_{os}	rig	86.89***	90.97***	81.81***	99.02***
	op	54.26**	61.20***	75.70***	93.04***
	GTU_{os}	47.05*	54.02**	49.51	65.29**
GTU_{od}	rig	68.10***	69.98***	76.72***	94.05***
	op	57.83***	60.79***	83.78***	100.90***
	GTU_{od}	41.59**	45.37**	56.63	73.21***
GTU_{oms}	rig	73.21***	76.64***	83.33***	100.47***
	op	50.51**	55.67***	78.86***	96.15***
	GTU_{oms}	41.63	46.87**	54.29	70.64**

Notes:
swald: Supremum Wald test
slr: Supremum likelihood ratio test

Each VAR equation for both rig count models is tested using swald and slr tests to examine the break in its relationship. The swald and slr tests are presented in Table 4.5. There are mixed rejections of the null hypothesis of no structural break, but the break in a relationship is mainly found when rig and oil price are the dependent variables. In the world rig count model, based on the swald test, t-statistics are rejected at a 5% significance level for all equations in JMU, GTU oil price, and GTU oil demand. A similar result is obtained in the slr test with t-statistics of GTU oil supply, and the GTU oil market-specific model is also rejected. For the rest of the OVX, GEPU, and GTU oil investment models, the equations with rig counts and oil price as dependent variables are mostly rejected, and equations with uncertainty indices as dependent variables are not rejected. Based on the swald test, JMU, OVX and GTU oil price equations are rejected for the North America rig count model. The rejection of the t-statistics is found when rig and oil price are the dependent variables. GEPU and most of the GTU indices as dependent variables are not rejected. Based on the slr test, all t-statistics in each equation are rejected, except the GEPU equation. Thus, a time-varying Granger Causality is considered because most short-run parameters' stability is rejected.

4.6.2 Time-varying Granger causality test

This section conducts time varying Granger causality tests to examine whether uncertainty helps to predict oil exploration activity and crude oil price in accommodating instability of the short-run parameters. The Granger causality test is carried out on VAR three-variable models consisting of the rig counts, oil price, as well as one of the following uncertainty indices: JMU macroeconomy uncertainty, OVX oil volatility uncertainty, GEPU economic policy uncertainty, and the five newly-proposed GTU indices (GTU oil price, GTU oil investment, GTU oil supply, GTU oil demand, and GTU oil market-specific). The Granger causality test includes a lag of uncertainty to understand whether past uncertainty values help forecast the current value of oil exploration activity and crude oil price.

Table 4.6 presents the results of testing time-varying Granger causality for uncertainty indices, rig counts, and oil price for the world and North America exploration models. The second column of Panel (a) and (b) in Table 4.6

describes that uncertainty Granger causes rig count as the null hypothesis of oil price does not Granger cause rig count is rejected. The third column shows that rig count Granger causes uncertainty in benchmark indices. Meanwhile, the causality from rig count to GTU indices is weak for the world exploration and is stronger from North America rig counts to all uncertainty indices. A similar direction of Granger causality is found between uncertainty and oil price.

Table 4.6: Time Varying Granger Causality test between uncertainty and global oil market

(a) World oil exploration activity model

Uncertainty indices[8]	H_0: Uncertainty does not Granger cause rig count	H_0: Rig count does not Granger cause uncertainty	H_0: Uncertainty does not Granger cause oil price	H_0: Oil price does not Granger cause uncertainty	H_0: Oil price does not Granger cause rig count	H_0: Rig count does not Granger cause oil price
JMU	29.298***	19.305***	89.943***	45.970***	41.755***	117.356***
OVX	48.407***	48.890***	57.953***	47.752***	69.116***	81.965***
GEPU	30.424***	39.526***	24.240***	35.951***	116.358***	48.543***
GTU_{op}	16.371***	11.037	12.457*	13.283**	56.462***	7.904
GTU_{oi}	23.276***	12.546*	33.227***	13.123*	60.252***	14.847**
GTU_{os}	23.156***	14.109*	32.188**	12.892*	70.796***	17.661***
GTU_{od}	22.347***	11.551*	11.327**	16.426*	48.054***	7.801
GTU_{oms}	22.683***	13.374*	31.101***	10.589*	53.074***	14.252**

(b) North America oil exploration activity model

Uncertainty indices[9]	H_0: Uncertainty does not Granger cause rig count	H_0: Rig count does not Granger cause uncertainty	H_0: Uncertainty does not Granger cause oil price	H_0: Oil price does not Granger cause uncertainty	H_0: Oil price does not Granger cause rig count	H_0: Rig count does not Granger cause oil price
JMU	257.286***	42.157***	215.662***	71.205***	263.899***	140.378***
OVX	127.196***	68.760***	195.167***	77.463***	176.894***	215.226***
GEPU	308.689***	345.789***	189.684***	283.607***	484.403***	175.689***
GTU_{op}	42.117***	17.240***	45.235***	28.948***	43.531***	30.487***
GTU_{oi}	141.388***	32.811***	48.463***	34.070***	77.163***	43.818***
GTU_{os}	108.215***	47.924***	77.221***	53.710***	63.003***	39.262***
GTU_{od}	100.945***	39.803**	88.078***	31.376***	78.165***	51.084***
GTU_{oms}	121.878***	38.007**	75.152***	38.432***	62.872***	46.795***

The uncertainty indices Granger cause oil price as shown in the fourth column of Table 4.6 as the null hypothesis of uncertainty does not Granger cause oil price is rejected for all equations. The fifth column indicates that oil price Granger causes uncertainty benchmark indices for the world exploration model. However, the causality from oil price to GTU indices in the world exploration model is weak. Oil price Granger causes all uncertainty indices for the North America exploration model. The last two columns show rig count

[8]Endogenous variables in VAR equations are world rig counts, crude oil price, and one of the uncertainty index; e.g. rig_w op OVX; rig_w op GTU_{oms}; etc.

[9]Endogenous variables in VAR equations are North America rig counts, crude oil price, and one of the uncertainty index; e.g. rig_{na} op OVX; rig_{na} op GTU_{oms}; etc.

and oil price causality. Oil price Granger causes rig count in both exploration models, and rig count Granger causes oil price in all uncertainty equations in the North America exploration model and most uncertainty equations in the world exploration model. A bidirectional causality is present in each null hypothesis in the exploration model of North America, which means the past value of uncertainty helps predict the current value of rig counts and oil price. The past value of rig counts is helpful in predicting the current value of uncertainty and oil price.

Overall, causality effects between various specifications of the GTU indices and the global oil market are consistent with benchmark measures and the existing studies. A strong causality from the GTU index to rig counts and oil price is expected. The causality effect from most of the uncertainty indices to the crude oil price shows that when public interest in some particular topics is intensified and there is uncertainty about the global oil market, it may change public decision-making in the oil market; thus, global oil supply and demand will be affected. When there are global supply and demand shocks, a consequence of this information gathering by a concerned public affects the investment insights and contributes to the fluctuating oil price, which leads to changes in oil exploration activity. Rig counts Granger cause uncertainty indices in both exploration models because the public will try to seek out information through an intensified web-based search when there is uncertainty in exploration.

The findings of the bidirectional causality between the GTU indices and crude oil price are expected and also support the finding of previous studies such as Guo and Ji (2013), Li et al. (2019), and Qadan and Nama (2018). Guo and Ji (2013) report that most of their Google Trends indices Granger cause Brent oil price. Qadan and Nama (2018) find that most of the investor sentiment indices, including Google Search Volume Index (GSVI), OVX, and Economics Policy Uncertainty (EPU) indices, significantly Granger-cause oil price returns using monthly data from 1986 to 2016. Their work also finds bidirectional causality between GSVI for ‘oil price’ and oil price return, which this thesis also confirms the results of bidirectional causality between GTU indices and oil price in both world and North America exploration models. Qadan and Nama (2018) argue that searching for information on the internet drives oil price volatility, and the volatility causes the investor to intensify their searching. Li et al. (2019) apply GSVI for various crude oil price markets. Their findings confirm a causality effect from GSVI to most of the spot crude oil prices (taken from the WTI, Brent, Daqing, and Dubai oil price markets). Their findings also argue that there is bidirectional causality from GSVI to WTI future crude oil return, but only unidirectional causality to other types of oil price markets.

The strong Granger causality from crude oil price to world and North America rig counts is economically plausible, as a high oil price will likely induce exploration activity. This finding also aligns with the literature that finds crude oil price shock significantly affects drilling activity (Toews and Naumov, 2015; Khalifa et al., 2017). The rig count is found to be Granger-causes oil price in all equations, except GTU oil price and GTU oil demand in the world exploration model.

The Granger causality findings thus confirm that the GTU index is one of the uncertainty measures that can help predict the changes in crude oil price and rig counts. As this study measures Google Trends-based uncertainty in a more comprehensive index that covers uncertainty in oil investment, demand, and supply, it improves on the previous studies that only apply basic search terms such as ’oil price’ or ’Brent’.

4.6.3 VAR and Impulse responses to rig counts, oil price, and uncertainty shocks

GTU index is constructed specifically for the global oil market uncertainty, and its construction incorporates the components of oil investment, oil supply, and oil demand. Other existing uncertainty measures are mainly based on macroeconomic and financial variables that are not directly associated with the oil market variables. The construction of the GTU index does not claim to be a better index but provides a new alternative public interest measure in the global oil market with a simple method compared to other uncertainty measures, which require computational intensive procedures. The VAR estimates are provided in Appendix B.3. and the predictive power of the econometric model is tested by calculating the RMSE obtained from the VAR estimates. Table 4.7 shows that the RMSE of benchmark and GTU indices are nearly similar. On average, the GTU indices have less error value than JMU and GEPU to predict rig counts and less than OVX and GEPU to predict oil prices.

Table 4.7: Root Mean Squared Error (RMSE) of the predicted rig counts, crude oil price, and uncertainty

(a) World oil exploration activity model

Predicted variable	VAR model							
	OVX	JMU	GEPU	GTU_{op}	GTU_{oi}	GTU_{os}	GTU_{od}	GTU_{oms}
Rig counts	0.0242	0.0252	0.0253	0.0249	0.0249	0.0245	0.0246	0.0248
Oil price	0.0925	0.0803	0.0897	0.0909	0.0878	0.0887	0.0914	0.0883
Uncertainty	0.1700	0.0059	0.1655	0.2526	0.1475	0.2089	0.1951	0.1732

(b) North America oil exploration activity model

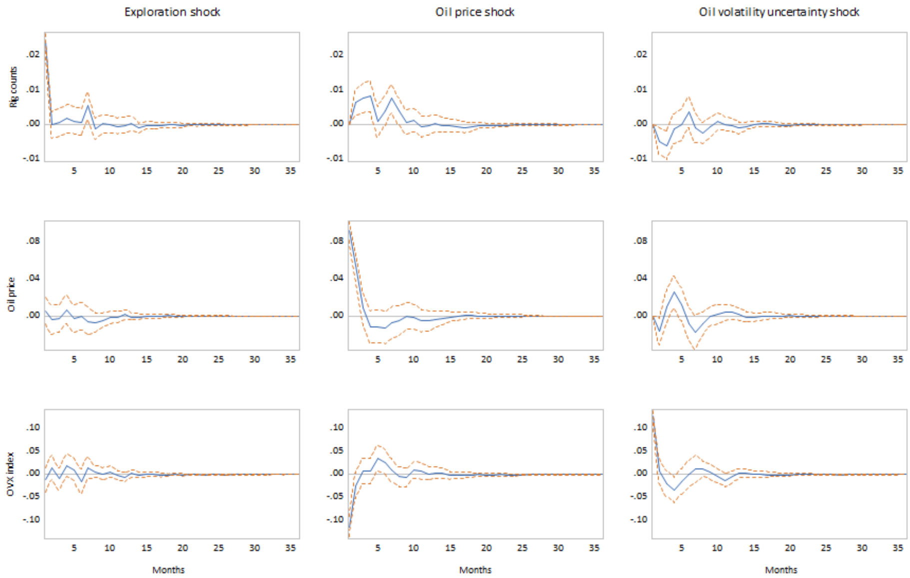
Predicted variable	VAR model							
	OVX	JMU	GEPU	GTU_{op}	GTU_{oi}	GTU_{os}	GTU_{od}	GTU_{oms}
Rig counts	0.0521	0.0769	0.0864	0.0642	0.0623	0.0637	0.0615	0.0617
Oil price	0.0935	0.0768	0.0899	0.0907	0.0893	0.0903	0.0917	0.0902
Uncertainty	0.1741	0.0056	0.1683	0.2529	0.1476	0.2131	0.1990	0.1770

Figures 4.17 and 4.18 show the point estimates of the impulse response functions in the forecasting horizon 36 months to a shock constructed using a Cholesky decomposition method. The impulse responses of a variable are

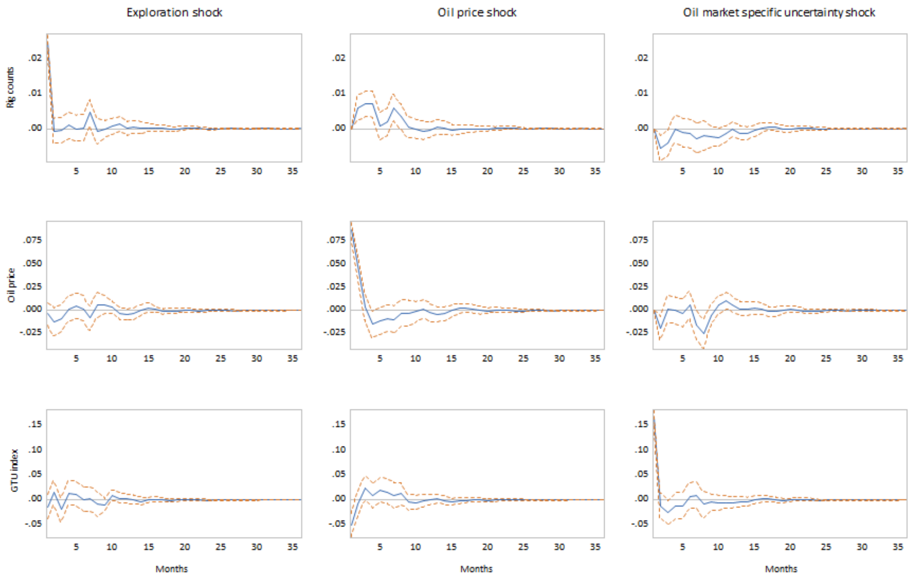
indicated on the left-hand side to shocks shown along the top of the figure in each row; for example, the impulse response of the OVX index to a one-standard-deviation world exploration shock is presented in the last row and the first column in Figure 4.17a. As for illustration, the OVX as a benchmark and a newly constructed GTU oil market specific indices are presented. Thus, the consistency of the model with the extant economic literature can be checked.

The first column of Figures 4.17 and 4.18 illustrate the responses of an endogenous variable to world and North America exploration shocks, respectively. The first row and first column show that unexpected exploration shocks cause a sharp decline in the rig counts upon impact. The declined effect exhibits a reversal between 2$^{\text{nd}}$ and 7$^{\text{th}}$ months in the world exploration model and between 4$^{\text{th}}$ and 13$^{\text{th}}$ months in the North America model. The North America rig counts respond more significantly to the exploration shocks over periods than the world.

The unexpected oil price shocks have positive and statistically significant effects on rig counts from 4$^{\text{th}}$ month in both world and North America, as depicted in the first row and second column of Figures 4.17a and 4.18a. The world rig counts responses to oil price shocks decay in the 20$^{\text{th}}$ month, while they are more persistent for North America rig counts. The unanticipated shock in crude oil price triggers a large, significant, immediate, and transitory response to the rig counts. With the prevalence of shale and tight oil exploration in North America, it would be of more benefit to carry out exploration activity when the crude oil price is high. A high oil price leads to the economic growth that will trigger a high demand for crude oil. Hence, the preference for embarking on an investment project with a view to maximising profit rather than postponing the venture. The significant response of exploration activity to oil price shock is also consistently described in Figure 4.3, which shows that as oil prices increase up to mid–2008, rig counts also rise sharply. Rig counts then decrease after the end of 2014 after a drop in crude oil price. Overall, the exploration



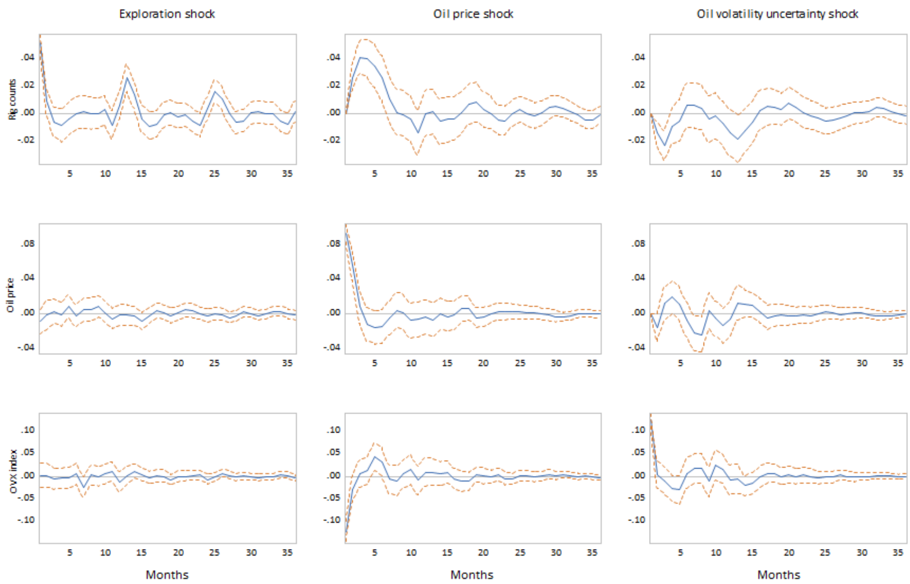
(a) OVX model



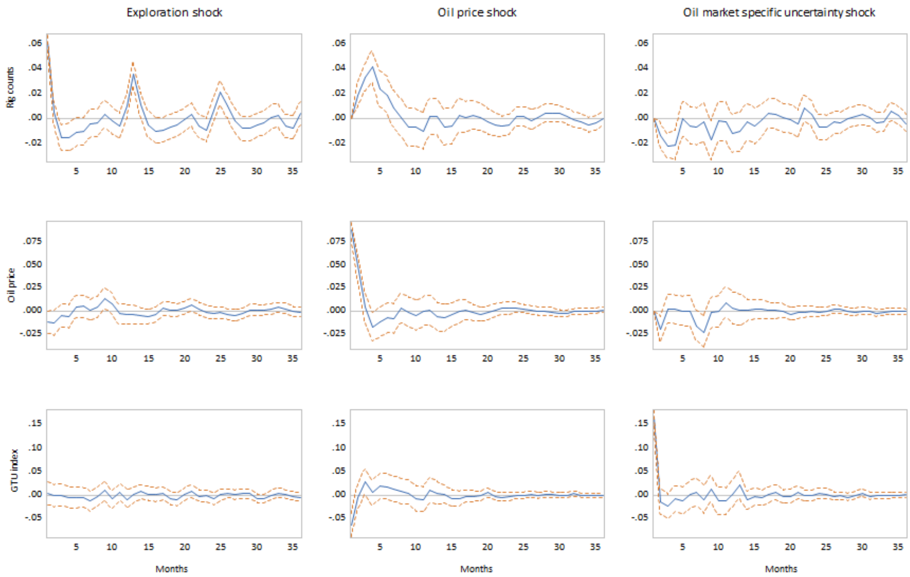
(b) GTU oil market specific model

Figure 4.17: Response to Cholesky one-standard deviation shock in world oil exploration model

activity response to an uncertainty shock indicates the expected sign after controlling the crude oil price variable in VAR. The significant positive response of rig counts to the crude oil price shock is also consistent with the economic literature, such as discussed in Kellogg (2014); Toews and Naumov (2015).



(a) OVX model



(b) GTU oil market specific model

Figure 4.18: Response to Cholesky one-standard deviation shock in North America oil exploration model

The first row and third column show that world rig counts respond negatively and statistically significant to the uncertainty shocks in 3rd month for OVX and 2nd month for GTU index, then decay after 19th month. The North America rig counts response to uncertainty shocks are significant on 3rd for OVX and 4th for GTU and are persistent.

The second row and first column depict that crude oil price responds negatively and is statistically significant to the world exploration shocks between 2nd and 7th month. An unexpected shock for North American exploration is only statistically significant on oil price after 9th month, and the effect is more persistent than to world exploration shock that decays after 26th month. The second column and second row indicate oil price shocks give a significant negative drop to the real oil price for up to four months, and the effects are dissolved after 28th month for the world exploration model, and it takes longer effect in North America exploration.

The third column and second row show that an unexpected oil market uncertainty shock is negative and significant on oil price between the 2nd and 8th months for both exploration models, then exhibits partial reversal after the 8th month. The crude oil price response to an unanticipated shock to the uncertainty indices is only transitory in both the world and the North American exploration models. The significantly large magnitude of the oil price decline in response to uncertainty shock implies that when the public gives high attention to certain dramatic events, this drives a significant reduction in investment. The higher the level of uncertainty, the slower the business cycle. As investment also plays an essential role in the business cycle, it eventually exerts a substantial effect on the crude oil price. It then leads to lower demand for crude oil and induces a decline in crude oil price. These findings are consistent with Qadan and Nama (2018), who employ other uncertainty indices and find a significant negative response of crude oil price to most of the financial uncertainty indices, such as investor sentiment index, consumer sentiment index, and financial stress index. The effect of the GTU shocks on oil exploration and crude oil price is expected and consistent with the benchmark models, i.e., the OVX, oil volatility index.

The impulse responses of GTU oil market-specific uncertainty to one standard deviation shock are presented in the last row in Figures 4.17 and 4.18. An unexpected oil exploration disruptions on the GTU index are negative and significant only in the 2$^{\text{nd}}$ period in response to the world exploration shock and insignificant to the North American exploration shock. The GTU oil market's specific uncertainty responses to an unexpected oil price shock are significant and negative in the first three months of both exploration models. Then, it exhibits a reversal to positive effect from the 3$^{\text{rd}}$ to 30$^{\text{th}}$ month for the world exploration model, and it takes more than 36 months for the uncertainty response to decay in the North America model. An unanticipated oil market-specific uncertainty shock causes a sharp decline in the GTU index up to 3$^{\text{rd}}$ period. The decline indicates a slight increase between the 3$^{\text{rd}}$ and 7$^{\text{th}}$ months for the world exploration model, and it takes a more prolonged increase up to the 13$^{\text{th}}$ month for the North America exploration model. The shock causes a gradual, cumulative decline in the GTU index following these increases and decay in 18$^{\text{th}}$ and 31$^{\text{st}}$ months for the world and North America models, respectively. Overall, impulse response functions of GTU oil market specific and oil market variables give similar characteristics of the impulse response by the OVX as an uncertainty benchmark index.

The negative relationship between uncertainty and crude oil price is consistent with the graphical interpretation shown in Figure 4.14a. The graph indicates that in most of the periods when the oil price drops, the GTU index rises. The GTU index reflects public interest in dramatic events. The cases of political conflict and pandemic lead to a low oil price while also triggering high public interest.

Robustness tests are carried out by reordering the variables into uncertainty index, rig count, and crude oil price, and also by using Brent price as a crude oil proxy. The impulse response graphs, as shown in Figures B.3-B.6 in the appendices B.5.1, are consistent with the main results in showing a significant negative response by exploration activities in both the world and North America to shocks in the GTU, as well as in the benchmark indices. The responses of crude oil prices with this reordering also show significant negative and transitory responses to the uncertainty shocks, as depicted in Figures B.5-B.6. Figures B.7a-B.9b explicitly show that applying Brent price also gives results that are consistent with the main findings, being a significant negative response of exploration and crude oil price to the uncertainty shocks, and a significant positive response of exploration activity to the crude oil price shock. The consistency of the GTU indices against the benchmark indices in impulse response functions proves the capability of the newly proposed GTU indices to measure uncertainty, particularly in the global oil market.

4.6.4 Forecast error variance decomposition of oil market to uncertainty shock

Forecast error variance decomposition (FEVD) decomposes the endogenous variables into the shocks of each VAR component. This section carries out variance decomposition to understand the extent to which uncertainty shock affects the variation in rig counts and crude oil price for three years ahead. Table 4.8 illustrates that, among the benchmark models, Jurado's macroeconomy uncertainty index contributes most to variance in the world and North America oil exploration activity (19.59% and 29.45%, respectively). This result implies that macroeconomy uncertainty is better able to forecast investment activity for a particular country than for the world as a whole.

The newly proposed GTU can explain the variance of world and North America exploration activity in the three-year ahead forecast within the range of 6-14%. GTU oil supply and GTU oil investment shocks contribute more than basic GTU oil price to explaining world oil exploration activity. GTU oil supply shock contributes largely to the variance of world oil rig counts (at 14.43%) while GTU oil investment shock contributes to the variance of North

America exploration (13.81%). All the newly proposed GTU indices contribute more than the basic GTU oil price to the North America rig counts.

Table 4.8: Forecast Error Variance Decomposition

(a) World exploration activity model

shocks[10]	rig	op
ε^{ovx}	9.40	2.68
ε^{jmu}	19.59	22.07
ε^{gepu}	1.96	3.68
$\varepsilon^{gtu_{op}}$	10.77	10.66
$\varepsilon^{gtu_{oi}}$	11.23	12.29
$\varepsilon^{gtu_{os}}$	14.43	11.08
$\varepsilon^{gtu_{od}}$	5.87	4.22
$\varepsilon^{gtu_{oms}}$	9.07	4.89

(b) North America exploration activity model

shocks[11]	rig	op
ε^{ovx}	15.72	12.15
ε^{jmu}	29.45	26.13
ε^{gepu}	3.00	4.33
$\varepsilon^{gtu_{op}}$	11.19	6.35
$\varepsilon^{gtu_{oi}}$	13.81	9.20
$\varepsilon^{gtu_{os}}$	12.56	6.68
$\varepsilon^{gtu_{od}}$	12.05	6.59
$\varepsilon^{gtu_{oms}}$	13.23	6.63

The contribution of the uncertainty benchmark indices to the crude oil price shows that Jurado's macroeconomy uncertainty index contributes most to the variance of crude oil price in world and North America exploration activities, at 22.07% and 26.13%, respectively. GTU indices contribute about 4-12% to the crude oil price in the world and North America exploration activity models. Among GTU shocks, GTU oil investment has the largest contribution to the crude oil price, being 12.29% and 9.20% in the world and North America exploration activities models, respectively.

Overall, the contribution of uncertainty shock to exploration activity in the

[10]The variation of the uncertainty shock to the world rig counts and crude oil price for 3-years ahead with the variables ordering; world rig counts, crude oil price, and one of the uncertainty index. For instance; variables ordering rig_w op ovx, the variation of ovx shock to world rig count is 9.4% and to oil price is 2.7%.

[11]The variation of the uncertainty shock to the North America rig counts and crude oil price for three-years ahead with the variables ordering; North America rig counts, crude oil price, and one of the uncertainty index. For instance; variables ordering rig_{na} op ovx, the variation of ovx shock to North America rig count is 4.8% and to oil price is 5.2%.

North America model exhibits a magnitude that is similar to its contribution to the world exploration model. The decomposition of the uncertainty shocks to oil exploration can be distinguished between the world and North America as the presence of shale oil exploration in North America causing less sensitivity to uncertainty shocks.

To summarise, first, the GTU indices for the oil market can, like the macroeconomy and financial market uncertainty measures, predict global oil market variables such as oil exploration and crude oil price. Second, the capability of the uncertainty indices to predict the variation of world exploration activity is similar to their capability to predict North America oil exploration activity. Third, the proposed GTU indices, particularly GTU oil investment, are better at explaining variation in the global oil market than the basic GTU oil price is. Therefore, the newly proposed GTU indices behave consistently with the existing benchmark indices and can be used as the new uncertainty indices for a global oil market. This offers a contribution to the scant literature on oil market uncertainty.

4.7 Conclusion

Prior work approaches oil market uncertainty indirectly, using financial and macroeconomic indicators rather than a direct measure specific to the oil market. This study proposes and constructs GTU, a concise index for the oil market based on Google Trends. The newly proposed GTU index directly measures public interest in oil investment, oil supply, and oil demand at the times when dramatic events have occurred and crude oil prices have fluctuated. The GTU index is flexible, free to access, and easy to construct compared to the existing benchmark, which requires heavy computation. It is based on the notion that public uncertainty is reflected in internet-search intensity; thus, increased searching about particular topics implies that greater public attention is being given to those topics.

This study contributes to the literature not only by constructing the new index, but also in presenting a comparison analysis against the benchmark indices. Further, it analyses how the uncertainty index affects oil exploration activity and the crude oil price. The findings of this study suggest that GTU indices exhibit consistent results that positively correlate with the benchmark indices, particularly with Jurado's macroeconomy and the oil volatility uncertainty indices. The GTU index captures three significant peaks (i.e., the Great Recession of 2008–09, the economic slowdown of 2015–16, and the Coronavirus pandemic of 2020). These are also captured by the benchmark macroeconomy and financial uncertainty indices, which increases confidence in the GTU index as a measure of oil market-specific uncertainty. The consistency of the results with the benchmarks instils confidence that the GTU index can be an alternative new index for measuring uncertainty via public attention, providing a bridge to the oil market variables.

A further contribution to the literature links crude oil price and oil exploration activity to the GTU index. Exploration plays an essential role in the oil market not only as a determinant of future production but also as a sentiment indicator that reflects investment activity in the oil market. This study distinguishes between world and North America exploration activities because the former is primarily associated with conventional oil exploration, while the latter is associated with non-conventional exploration. Thus, the GTU oil investment index can go further than the basic GTU oil price index in explaining variance in exploration activity and oil prices. Although the oil investment index shows major peak points that are similar to those of the basic index, it exhibits more distinguished and significant peak points that provide more information that nuances the explanation of exploration activity and oil price. For instance, in the period when rig counts decline and there is a drop in the crude oil price (i.e., between the end of 2014 and early 2016), GTU oil investment shows a more significant upward trend than does the basic GTU oil price. On the other hand, when rig counts and crude oil price rise (between mid–2005 and early 2008, and from early 2010 to mid–2014), the GTU index exhibits a decline trend that is steeper than that of the flatter GTU oil price.

Unanticipated GTU shocks cause significant and negative responses in oil exploration activity and crude oil price. Distinguishing world oil exploration from North American oil exploration allows for a more straightforward interpretation of this by showing that uncertainty affects North America exploration activity as significantly as it affects world exploration activity. As high uncertainty leads to economic slowdown, this also causes a low demand for crude oil. In consequence, crude oil price also declines significantly in response to the high uncertainty. An unanticipated shock to the crude oil price exerts a strongly significant positive effect on the exploration activities in both the world and North America; this is consistent with the theoretical perspective that a high oil price triggers more investment activity as producers seek to gain more profit rather than take the risk of postponing it. A robustness test that reorders the variables and applies Brent prices gives results that are consistent with those of the base model.

GTU shock can explain the variance in exploration activity and crude oil price with the same degree as the macroeconomy and financial market uncertainty benchmark indices do. In the models for world and North America oil exploration activities, GTU oil investment contributes more to the oil market variables than other specification. As the new measure of uncertainty, GTU index behaves as a conduit between public attention about dramatic events and the global oil market variables. The GTU index is useful for measuring uncertainty that has an impact on exploration activity and crude oil price fluctuation. Future research could investigate the ability of the GTU oil investment index to affect macroeconomic indicators. Furthermore, applying the GTU oil market-specific index to the oil-exporting and oil-importing countries could be an interesting task for future studies.

Chapter 5

Conclusion

The fluctuations in oil price since the 1970s has prompted a number of studies to discuss the relationship between oil price and the macroeconomy, mainly focusing on the US. Indeed, ever since crude oil became a globally-traded commodity, dramatic changes in its price have impacted on the economy and the financial markets, triggering ongoing debate as to the drivers of the enormous price fluctuations. Most researchers agree that they are driven by the interaction of supply and demand, however this is only the case in the short run. In the long run, demand is determined by global economic activity, while supply is driven by exploration activity in the past. While there is some literature on this, there are not many empirical studies. In addition, supply disruptions such as the Suez Crisis (January 1957–February 1957), the OPEC embargo (November 1973–February 1974), and the first Persian Gulf War (August 1990–October 1990) have caused oil prices to increase. The crude oil market is thus often considered to be a political market because geopolitical developments matter to it. This also explains why the term 'uncertainty' is often used in discussions about the crude oil market.

Understanding the feedback loop between oil supply, demand, uncertainty, and price is important to understand the direction that the future oil market might take. This thesis sheds light on that relationship, not only in the short run but also in the long run. It thus fills a gap in the existing literature, which is mainly concerned with the short-term supply and demand, and also focuses on oil production rather than exploration. Through Chapter 2's study on exploration activity simulation, this thesis takes into account the role of exploration in determining oil supply in the future. The economic variable, crude oil price, is also accounted for in the empirical analysis to understand how much its movements affect exploration activity. Finally, the thesis takes into consideration the fact that it takes a few years for actual oil production to commence once the exploration phase has been completed.

Turning to the impact of crude oil price shocks on supply and demand, most of the literature assumes that spikes and drops in the crude oil price have symmetrical effects on supply and demand. Chapter 3 of this thesis discusses the supply side, captured by world oil production, and reveals an important insight into whether positive or negative oil price shocks have a dominant impact on supply and demand. This contributes to the current literature, as does the chapter's exploration of the long-run relationship between supply, demand, and oil price. This addresses a shortcoming in the literature, which focuses more on the supply role in the short term.

Supply and demand are clearly fundamental to setting the price of oil, but the term 'uncertainty' is also closely linked to the crude oil price. Much of the literature allows the concept of uncertainty into the global oil market model as a means of capturing unexpected factors that contribute to oil price fluctuation and which supply and demand cannot explain. The literature defines uncertainty as market expectation, public interest, or sentiment indicators about future oil prices. However, it is challenging to measure uncertainty through direct observation, and so the current studies approach uncertainty through forecast-based oil price volatility or forecast-based macroeconomic indicators. There is therefore no uncertainty index specific to the global oil market that links uncertainty to oil price, supply, and demand; this is a knowledge gap that Chapter 4 of this thesis aims to fill.

In response to gaps in the existing literature, Chapter 2 attempts to capture exploratory efficiency as a result of exploratory effort. It does this through the simulation case of a mature petroleum province, the Norwegian Continental Shelf, and investigates empirically how this is affected by the crude oil price fluctuation. A Monte Carlo simulation is applied to generate the discovery size and time between discoveries sequence. The benefit of applying Monte Carlo simulation is that it does not require extensive numerical analysis and can generate a simulation resembling a complex system. The simulation results find, with high confidence, that the average time between one and the next successful oil discovery is less than ten months, with the average discovery size being smaller than the past giant discoveries. The key findings from the simulation exercises are twofold; first, the intervening time between oil discoveries gets shorter as they occur more frequently, capturing more frequent exploratory effort. Second, on average, larger size discoveries are made with less frequency than the smaller size. Hence, the most common pattern for oil discovery is for smaller discoveries to be made more frequently and with a shorter time between one and the next discoveries.

The effect of crude oil price movements on exploratory efficiency is also analysed in Chapter 2. A long-run relationship is found between exploratory effort, efficiency, and crude oil price. Adding a structural break during Global Financial Crisis in 2008 confirms the presence of the long-run equilibrium relationship. The empirical analysis concludes that high crude oil price is associated with more exploration wells and a shorter time between two discoveries in the long run. A high oil price incentivises oil producers to drill more often. A high oil price is also associated with a larger discovery size. The result supports the argument by Mohn (2008) that the positive relationship between oil price and size is because high oil price gives an incentive to explore frontier areas where there is the prospect of making a larger size discovery than is available in a mature field.

The findings from Chapter 2 contribute to the academic literature by developing a toolbox for Monte Carlo simulation that can be applied to other petroleum provinces. Further, the base case scenario can use the findings from this study; for instance, the insight that there is 90% probability of average waiting time being shorter than 316 days with a discovery size of less than 94 million standard cubic metres of oil equivalents. The simulation also identifies the last discovery date by reference to the discovery rate (i.e., the last discovery to be earlier than 2089 for the base case parameter). It allows the researcher to capture the result of exploration activity in various petroleum regions by simply varying the parameters. Second, the distribution profile of the oil discovery can be adjusted according to the profile of the focal petroleum region. This piece of information is particularly useful for a preliminary study of exploration activity in new frontier areas.

Chapter 3 provides insights into the broader scope of world oil production (representing global oil supply) and its interaction with the crude oil price and global demand. The chapter emphasises the role of supply in the long term, asymmetric effects, and structural break, something that have been underestimated in the existing literature. The empirical framework is carefully chosen to accommodate the mixed stationary and non-stationary variables in levels. The empirical findings from the basic Autoregressive Distributed Lags (ARDL) model are that a long-run relationship is present between world oil supply, global demand, and crude oil price. Incorporating the estimated break in January 2009 confirms the long-run equilibrium relationship between crude oil prices, supply, and demand. Oil price relates negatively with oil production, and positively with global demand in the long run, while global demand relates positively with world oil production. The common view that supply does not respond to the oil price shock is more relevant to the short run supply, given that oil production is not very flexible. The finding indicates that a change in demand is associated with a greater change in oil price, while a change in supply is less so. Further, an increase in oil price is associated with a slight reduction in world oil production, whereas its association with demand is of greater magnitude. Oil price will make a large adjustment when the oil supply, demand, and oil price deviate from the level in long-run equilibrium.

Consistent with the basic ARDL model, further analysis via an asymmetric ARDL model shows that the long-run equilibrium between world oil supply, global demand, and oil price is significant. Three key findings are made in the asymmetric analysis. First, supply and demand indicate a stronger link to positive oil price shocks. Second, crude oil price relates negatively to positive and negative supply shocks, and positively to positive and negative demand shocks. Third, oil prices have a stronger link to positive and negative supply shocks than to positive and negative demand shocks. There is no substantial difference in the accumulated impacts of positive and negative changes in demand on the oil price. However, the oil price impacts appear to be greater for negative change in demand than for positive change in demand in the short run. After considering a structural break in the asymmetric ARDL model, the results confirm that the interaction of the break with positive and negative global demand shocks is significant on the crude oil prices. The break interaction with the negative supply shock affects oil prices more strongly than the positive supply shock. Both positive and negative supply shocks strongly affect demand, whilst the oil supply does not respond to the structural break in the short run.

Incorporating uncertainty in the global oil market is crucial to capturing unprecedented events that fundamental supply and demand cannot explain. Chapter 4 proposes an uncertainty index based on Google Trends (GTU), which captures public interest in oil price, investment, supply, and demand. The GTU indices measure uncertainty through directly observable data. This contrasts with the existing indices, which are mainly derived from forecast-based macroeconomic variables. The newly proposed GTU index draws on individuals' use of a web-based search engine to seek out information. This behaviour is intensified when there is uncertainty. High public interest in a topic is captured by the increased frequency with which individuals use the internet to 'Google' specific information, using keywords related to the topic. Uncertainty about a specific topic is therefore represented by high relative frequency volume of the topic's search terms in Google Trends.

The newly proposed GTU index can be used as a alternative index for measuring uncertainty in the global oil market and for predicting oil price fluctuation. When the GTU indices are compared with the benchmark uncertainty indices, there is a positive correlation between the GTU index and the benchmark indices, particularly the oil volatility (OVX) and Macroeconomy Uncertainty (JMU) indices proposed by Jurado et al. (2015). Furthermore, empirical analysis examining the relationship between uncertainty, exploration, and crude oil price indicates that the impulse response functions of the GTU index also illustrate consistent results with the benchmark indices.

From the findings from Chapter 4, it may be concluded that the GTU indices behave as a transmission channel of market concern about the unexpected worldwide events that significantly impact on oil price. Using rig counts as the exploration activity proxy and disaggregating these counts into world and North America rig counts (to account for the dominant non-conventional shale oil in North America), there is a significant negative relationship between uncertainty and oil price in the short term. The explanation of the interaction between uncertainty, oil price, and exploration is that high uncertainty causes a slowdown in the economy. The economic downturn then triggers a lower demand in business sectors, leading to a drop in oil prices. A negative relationship is also found between uncertainty and exploration activity. A project's cash flow would be affected by high uncertainty, leading to a decision to delay or cancel the project. As a consequence, a decline occurred in the upstream investment. The empirical findings also indicate that exploration activity relates positively and significantly to oil price shocks for all uncertainty index models. The empirical findings from Chapter 4 confirm the importance of crude oil price on the exploratory efforts, which was suggested in Chapter 2.

The future direction of the oil market is associated with the role crude oil plays in the energy transition. The energy transition initiatives and the uncertainty in future demand have thus become a crucial issue for the oil-producing countries, who must decide whether to keep finding new reserves and producing oil to fill the supply gap, or to leave the oil underground. From the supply side, a shift to clean energy by the industrial sectors and strong government policy towards energy transition will slow down the world oil production growth.

Current global attention is concerned with the peak oil. Historically, peak oil was perceived as a supply issue in a theory going back to Hubbert (1962), who argues that production peak follows a bell-shaped curve. The notion is that the once the peak has been reached, the production rate will decline whilst demand continues to rise. Oil prices have historically been strongly associated with concerns about peak oil supply. However, this theory is now open to criticism since it underestimated the additional reserve growth from non-conventional oil production. Thus, as a result of the development of non-conventional oil production and the strong global economy post–2000, concern about peak oil switched to the issue of peak oil demand. Crude oil price and demand shock are strongly influenced by each other, as evidenced by the extreme oil pricing episode of the Global Financial Crisis and, more recently, the Coronavirus pandemic of 2020 which caused a collapse in global demand, hitting the oil price in consequence. Indeed, in April 2020, the monthly average real price of crude oil reached its lowest level of 18 USD per barrel. This negative demand shock also had the consequence of delaying energy projects and reducing spending in upstream investment (International Energy Agency, 2022). Post-pandemic and in the medium term, the International Energy Agency (2022) forecasts that the emerging economies (most particularly China, India, and certain other Asian countries) will contribute to more than 90% of the growth in global oil demand between 2019 and 2026. On the demand side, a rebound in oil demand after the 2020 pandemic has prompted an increase in world oil production to meet it.

The analysis of the world oil supply, demand, uncertainty, and crude oil price shocks in this thesis provide an academic exercise that is related to the role of crude oil in the energy transition. The result from Chapter 2 indicates that a reduction in exploration activity in the mature basin will cause a longer time between oil discoveries. In addition, a high oil price increases exploration activity and reduces the time between discoveries. The toolbox for the simulation of the variable of interest in exploration can usefully be applied to other petroleum provinces to illustrate the average time between oil discoveries and the discovery size. It is necessary to take such findings into account when analysing supply forecast as a factor contributing to a slowdown in the global oil supply. Through the sequence of discoveries, the simulation exercise also helps illustrate how much reserve growth is required to fill the supply gaps. McKinsey & Company (2021) reports that oil production needs to add 38 million barrels per day if it is to meet demand by 2040, and that even if the energy transition scenario is accelerated, new oil production of 23 million barrels per day will still be required to meet global demand. Thus, policy measures are required to boost exploration and production activities in the frontier petroleum provinces, leading to a shorter time between discoveries and additional reserve growth to fill the gaps in meeting demand.

The presence of a long-run equilibrium between global oil supply, demand, and crude oil price, and the asymmetric effect disclosed in Chapter 3 are useful for illustrating the long-term oil market outlook. The asymmetric analysis reveals two things. First, both supply and demand are more strongly linked to positive oil price shock than to negative oil price shock. Second, demand and oil price are more strongly linked to positive supply shock than to negative supply shock. However, the elasticity of long-run price to supply is very small, while global demand change is associated with a higher oil price. In the medium term, the International Energy Agency (2022) forecasts that there will be a strong rise in demand up to 2026, sparked by a recovery in post-pandemic economic activity and by rising populations and incomes in Asian countries. The finding of Chapter 3 provides an insight that a positive demand shock is associated with an increase in oil price. Another spike in oil price is then expected during the global demand rebound. Linking this finding with an energy transition scenario characterised by more stringent policy measures, a slowdown in demand is expected in the longer term. Thus, it is also expected that crude oil prices will experience another period of low episodes following this negative demand shock.

The asymmetric findings can have some relevant policy implications. First, a policy measure that is conducive to an acceleration of the exploration and production phases is required, particularly to overcome the high oil price scenario. The implementation of a stable policy is not only necessary for supporting existing projects but is also required to boost investment if technological progress is to be achieved. Second, national governments should design policy that can respond to positive or negative oil price shocks, particularly when these hit during a period of economic stagnation or recession. Although the short-run effects of oil price shocks are less significant than the long-run effects, policymakers must be aware of how a sudden enormous shortfall in world oil supply or global demand contribute to the crude oil price fluctuation. At the very least, business portfolio diversification is required to reduce oil sector dependency, especially in the resource-rich countries.

The discussion of uncertainty in Chapter 4 focuses on the measurement of public interest in the global oil market. The key finding is that the GTU index captures the intensification of public interest when there is uncertainty due to unexpected circumstances. Public interest becomes particularly intense when the oil price declines. The International Energy Agency (2022) reports that there have been a few reasons for the increase in uncertainty about oil demand. These are related to oil exploration and oil production projects: cancellation, delay, low number of projects approved in the pandemic era, low oil price, and government policy to speed up the deployment of the energy transition scenario. The newly proposed GTU index provides a new measure that captures public interest in the oil market, which can be linked to extreme oil price movements, particularly when crude oil prices are low.

There are also policy implications from the findings from Chapter 4. Policymakers must anticipate market behaviour, as when there are disruptions, the public tends to behave irrationally. Without an immediate response from government, public reaction to disruption will deteriorate the economy and contribute to the crude oil price shock. Second, the policymakers can incorporate the uncertainty component into their sensitivity analysis and use the findings from the base case scenario in the oil market model to make a better strategy for investing in exploration activity. GTU oil investment and GTU oil supply have higher predictability for the world and North America exploration models; thus, these indices can be used as a proxy in extending research on crude oil production. Overall, this thesis illustrates the importance of a feedback loop between supply, demand, uncertainty, and oil price, which is relevant to further research on the role of oil in the energy transition scenario.

Appendix A

Exploration activity and the crude oil price: evidence from the mature petroleum province

A.1 R code for Norwegian oil discovery simulation

```
\begin{singlespace}
#NPD data
R_npd_disc_2 <- read_excel("PHD/ABERDEEN_1stYear/
Edited␣data/R_npd_disc_2.xlsx")
View(R_npd_disc_2)
year <- R_npd_disc_2$year
seq <- R_npd_disc_2$seq
recov <- R_npd_disc_2$recov
lrecov <- log(recov)
#Data and descriptive statistics
#Packages for date and data table formatting
```

```
library(data.table)
library(dplyr)
library(lubridate)
year <- ymd(as.Date(year))

#Plot of discovery size
par(mfrow=c(1,1))
plot(year,recov,xlim=as.Date(c("1967-07-09","2024-12-31")),
ylim=c(0,800),type="line",col="red",xlab="Year",
ylab="Discovery␣size␣[mill␣Sm3]")
par(mfrow=c(2,1))
plot(year,recov,xlim=as.Date(c("1967-07-09","2024-12-31")),
ylim=c(0,800),type="line",col="red")
plot(year,lrecov,xlim=as.Date(c("1967-07-09","2024-12-31")),
ylim=c(0,800),type="line",col="red")

#Fit distribution of NPD data
library(fitdistrplus)
#Discovery size
plot(recov, pch=20)
plotdist(recov, histo = TRUE, demp = TRUE)
descdist(recov, discrete=FALSE, boot=500)
#summary statistics
#------
#min: 0.02  max: 578.69
#median: 7.9
#mean: 41.08057
#estimated sd: 95.87531
#estimated skewness: 3.78726
#estimated kurtosis: 18.26716
summary(recov)
```

```
# Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#   0.02    2.29    7.90   41.08   24.10  578.69
fit_w  <- fitdist(recov, "weibull")
fit_g  <- fitdist(recov, "gamma")
fit_ln <- fitdist(recov, "lnorm")
fit_e <-  fitdist(recov, "exp")

gofstat(list(fit_w, fit_g, fit_ln, fit_e), fitnames =
c("weibull",
"gamma", "lnorm", "exp"))
#Goodness-of-fit statistics
#                    weibull     gamma      lnorm         exp
#Kolmogorov-Smirnov statistic 0.1003794 0.1674752 0.03934430
# 0.3846621
#Cramer-von Mises statistic   0.2822424 1.0168108 0.02963787
# 7.9361813
#Anderson-Darling statistic   1.7561936 5.2021768 0.19628837
# 52.3348718
#
#Goodness-of-fit criteria
#                    weibull    gamma    lnorm      exp
#Akaike's Information Criterion 1174.196 1198.729 1158.183
# 1331.781
#Bayesian Information Criterion 1180.094 1204.627 1164.081
# 1334.730
summary(fit_ln)
#Fitting of the distribution ' lnorm ' by maximum likelihood
#Parameters :
#        estimate Std. Error
#meanlog 1.980080  0.1685455
#sdlog   2.001367  0.1191795
```

```
#Loglikelihood:  -577.0916   AIC:  1158.183   BIC:  1164.081
#Correlation matrix:
#              meanlog        sdlog
#meanlog 1.000000e+00 2.854558e-10
#sdlog   2.854558e-10 1.000000e+00

par(mfrow=c(2,2))
plot.legend <- c("Weibull","gamma","lognormal","exponential")
denscomp(list(fit_w, fit_g, fit_ln, fit_e),
legendtext = plot.legend)
cdfcomp (list(fit_w, fit_g, fit_ln, fit_e),
legendtext = plot.legend)
qqcomp  (list(fit_w, fit_g, fit_ln, fit_e),
legendtext = plot.legend)
ppcomp  (list(fit_w, fit_g, fit_ln, fit_e),
legendtext = plot.legend)

# Waiting time
for (i in 1:141){
deltat <- year - lag(year)
}
deltat <- as.numeric(deltat)
View(deltat)
par(mfrow=c(1,1))
plot(deltat,type="line",col="dark blue",xlab="Sequence
of discovery",
ylab="Waiting time (days)")
t <- deltat[2:141]
df.t <- data.frame(t)
library("writexl")
write_xlsx(df.t,"C:\\Users\\estit\\Documents\\PHD\\
```

```
ABERDEEN_1stYear\\R\\t.xlsx")
plot(t, pch=20)
plotdist(t, histo = TRUE, demp = TRUE)
descdist(t, discrete=FALSE, boot=500)
#summary statistics
#------
#min:  0   max:  899
#median:  88
#mean:  136.2357
#estimated sd:  157.707
#estimated skewness:  2.375989
#estimated kurtosis:  9.74944
summary(t)
#Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
#   0.00   37.75   88.00  136.24  176.25  899.00
summary(year)
# Min.       1st Qu.       Median      Mean      3rd Qu.
# "1967-07-09" "1983-12-25" "1998-05-03" "1997-08-20"
# "2011-10-27"
#        Max.
# "2019-09-27"
fit_w_t  <- fitdist(t[t>0], "weibull")
fit_g_t  <- fitdist(t[t>0], "gamma")
fit_ln_t  <- fitdist(t[t>0], "lnorm")
fit_exp_t <- fitdist(t[t>0], "exp")
gofstat(list(fit_w_t, fit_g_t, fit_ln_t, fit_exp_t),
fitnames = c("Weibull", "gamma", "lognormal", "exponential"))
#Goodness-of-fit statistics
#                           Weibull      gamma   lognormal
# exponential
#Kolmogorov-Smirnov statistic 0.04587222 0.05537463 0.06341836
```

```
# 0.06654069
#Cramer-von Mises statistic   0.05189915 0.07795903 0.12572442
# 0.11642037
#Anderson-Darling statistic   0.35880368 0.47135138 0.89568718
# 0.65373694
#
#Goodness-of-fit criteria
#                            Weibull    gamma lognormal
# exponential
#Akaike's Information Criterion 1649.132 1649.838  1653.161
# 1648.192
#Bayesian Information Criterion 1655.001 1655.707  1659.030
# 1651.127
summary(fit_exp_t)
#Fitting of the distribution ' exp ' by maximum likelihood
#Parameters :
#        estimate   Std. Error
#rate 0.007287789 0.0006062372
#Loglikelihood:  -823.0962   AIC:  1648.192   BIC:  1651.127

par(mfrow=c(2,2))
plot.legend <- c("Weibull","gamma","lognormal","exponential")
denscomp(list(fit_w_t, fit_g_t, fit_ln_t, fit_exp_t),
legendtext = plot.legend)
cdfcomp (list(fit_w_t, fit_g_t, fit_ln_t, fit_exp_t),
legendtext = plot.legend)
qqcomp  (list(fit_w_t, fit_g_t, fit_ln_t, fit_exp_t),
legendtext = plot.legend)
ppcomp  (list(fit_w_t, fit_g_t, fit_ln_t, fit_exp_t),
legendtext = plot.legend)
```

```
# Simulation of oil discovery
# Simulated 141 discovery size (10,000 replication) -
# NCS parameters
set.seed(93)
sim.ln.141 <- replicate(n=10000, expr=rlnorm(n=141,
meanlog=1.980080, sdlog=2.001367))
par(mfrow=c(1,1))
hist(sim.ln.141)
df.sim.ln.141 <- data.frame(sim.ln.141)
library("writexl")
write_xlsx(df.sim.ln.141,"C:\\Users\\estit\\Documents\\PHD\\
ABERDEEN_1stYear\\R\\sim.ln.141.xlsx")

# KS test simulated and actual discovery size distribution
ks.test(recov,sim.ln.141[,1],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  recov and sim.ln.141[, 1]
#D = 0.12057, p-value = 0.257
#alternative hypothesis: two-sided
ks.test(recov,sim.ln.141[,100],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  recov and sim.ln.141[, 100]
#D = 0.049645, p-value = 0.995
#alternative hypothesis: two-sided
ks.test(recov,sim.ln.141[,500],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  recov and sim.ln.141[, 500]
```

```
#D = 0.11348, p-value = 0.3241
#alternative hypothesis: two-sided
ks.test(recov,sim.ln.141[,1000],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  recov and sim.ln.141[, 1000]
#D = 0.056738, p-value = 0.9771
#alternative hypothesis: two-sided
ks.test(recov,sim.ln.141[,5000],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  recov and sim.ln.141[, 5000]
#D = 0.085106, p-value = 0.6868
#alternative hypothesis: two-sided
ks.test(recov,sim.ln.141[,10000],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  recov and sim.ln.141[, 10000]
#D = 0.06383, p-value = 0.9362
#alternative hypothesis: two-sided

#Quantile
install.packages("matrixStats")
require(matrixStats)
library(matrixStats)
probs <- c(0.10, 0.25, 0.375, 0.5, 0.625, 0.75, 0.90)
quant.ln141 <- matrix(sim.ln.141, nrow = 141, ncol = 10000)
q.ln141 <- rowQuantiles(quant.ln141, probs = probs)
View(q.ln141)
df.q.ln141 <- data.frame(q.ln141)
write_xlsx(df.q.ln141,"C:\\Users\\estit\\Documents\\PHD\\
```

```
ABERDEEN_1stYear\\R\\quantile_ln141.xlsx")

#MAPE test for simulated disc size
install.packages("Metrics")
require(Metrics)
library(Metrics)
mape(recov,sim.ln.141[,1])
# [1] 53.73598
mape(recov,sim.ln.141[,100])
#  [1] 67.74372
mape(recov,sim.ln.141[,500])
# [1] 75.28714
mape(recov,sim.ln.141[,1000])
# [1] 19.56513
mape(recov,sim.ln.141[,5000])
# [1] 38.93405
mape(recov,sim.ln.141[,10000])
# [1] 67.87165

# Changing parameters (meanlog and sdlog)
# low mean and sd
set.seed(932)
ln.l <- replicate(n=10000, expr=rlnorm(n=141, meanlog=1.7675,
sdlog=1.5))
par(mfrow=c(1,1))
hist(ln.l)
df.ln.l <- data.frame(ln.l)
write_xlsx(df.ln.l,"C:\\Users\\estit\\Documents\\PHD\\
ABERDEEN_1stYear\\R\\ln.l.xlsx")

# KS test simulated and actual discovery size distribution
```

```
ks.test(recov,ln.l[,1],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  recov and ln.l[, 1]
#D = 0.13475, p-value = 0.1545
#alternative hypothesis: two-sided
ks.test(recov,ln.l[,100],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  recov and ln.l[, 100]
#D = 0.14184, p-value = 0.1172
#alternative hypothesis: two-sided
ks.test(recov,ln.l[,500],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  recov and ln.l[, 500]
#D = 0.13475, p-value = 0.1545
#alternative hypothesis: two-sided
ks.test(recov,ln.l[,1000],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  recov and ln.l[, 1000]
#D = 0.13475, p-value = 0.1545
#alternative hypothesis: two-sided
ks.test(recov,ln.l[,5000],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  recov and ln.l[, 5000]
#D = 0.14894, p-value = 0.08763
#alternative hypothesis: two-sided
ks.test(recov,ln.l[,10000],alternative="two.sided")
```

```
#Two-sample Kolmogorov-Smirnov test
#
#data:  recov and ln.l[, 10000]
#D = 0.15603, p-value = 0.0646
#alternative hypothesis: two-sided

#Quantile
quant.lnl <- matrix(ln.l, nrow = 141, ncol = 10000)
q.lnl <- rowQuantiles(quant.lnl, probs = probs)
View(q.lnl)
df.q.lnl <- data.frame(q.lnl)
write_xlsx(df.q.lnl,"C:\\Users\\estit\\Documents\\PHD\\
ABERDEEN_1stYear\\R\\quantile_lnl.xlsx")

#MAPE test for simulated disc size
mape(recov,ln.l[,1])
#[1] 11.08655
mape(recov,ln.l[,100])
#[1] 9.591367
mape(recov,ln.l[,500])
# [1] 12.7842
mape(recov,ln.l[,1000])
# [1] 13.98826
mape(recov,ln.l[,5000])
# [1] 9.406069
mape(recov,ln.l[,10000])
# [1] 46.04467

# high mean and sd
set.seed(933)
ln.h <- replicate(n=10000, expr=rlnorm(n=141, meanlog=2.3575,
```

```
sdlog=2.1500))
par(mfrow=c(1,1))
hist(ln.h)
df.ln.h <- data.frame(ln.h)
write_xlsx(df.ln.h,"C:\\Users\\estit\\Documents\\PHD\\
ABERDEEN_1stYear\\R\\ln.h.xlsx")

# KS test simulated and actual discovery size distribution
ks.test(recov,ln.h[,1],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  recov and ln.h[, 1]
#D = 0.099291, p-value = 0.4904
#alternative hypothesis: two-sided
ks.test(recov,ln.h[,100],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  recov and ln.h[, 100]
#D = 0.15603, p-value = 0.0646
#alternative hypothesis: two-sided
ks.test(recov,ln.h[,500],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  recov and ln.h[, 500]
#D = 0.14894, p-value = 0.08763
#alternative hypothesis: two-sided
ks.test(recov,ln.h[,1000],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  recov and ln.h[, 1000]
#D = 0.078014, p-value = 0.7842
```

```
#alternative hypothesis: two-sided
ks.test(recov,ln.h[,5000],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  recov and ln.h[, 5000]
#D = 0.14894, p-value = 0.08763
#alternative hypothesis: two-sided
ks.test(recov,ln.h[,10000],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  recov and ln.h[, 10000]
#D = 0.12057, p-value = 0.257
#alternative hypothesis: two-sided

#Quantile
quant.lnh <- matrix(ln.h, nrow = 141, ncol = 10000)
q.lnh <- rowQuantiles(quant.lnh, probs = probs)
View(q.lnh)
df.q.lnh <- data.frame(q.lnh)
write_xlsx(df.q.lnh,"C:\\Users\\estit\\Documents\\PHD\\
ABERDEEN_1stYear\\R\\quantile_lnh.xlsx")

#MAPE test for simulated disc size
mape(recov,ln.h[,1])
# [1] 517.3936
mape(recov,ln.h[,100])
# [1] 43.27765
mape(recov,ln.h[,500])
# [1] 140.5207
mape(recov,ln.h[,1000])
# [1] 290.6018
```

```
mape(recov,ln.h[,5000])
# [1] 356.3883
mape(recov,ln.h[,10000])
# [1] 126.5435

# Density of various discovery size parameter
par(mfrow=c(1,1))
plot(density(ln.l), main="Density␣simulated␣discovery␣size
(Lognormal␣distribution)",col="blue",xlim=c(110,3000),
ylim=c(0,0.32))
lines(density(sim.ln.141),col = "red")
lines(density(ln.h),col = "orange")
legend("topright", c("mu=1.7675␣s=1.50","mu=1.9801␣s=2.00",
"mu=2.3575␣s=2.15"),col = c("blue","red","orange"), lty = 1)

# Simulated waiting time (10,000 replication) - NCS parameters
set.seed(411)
sim.t.141 <- replicate(n = 10000, expr = rexp(n = 141,
rate = 0.007287789))
par(mfrow=c(1,1))
hist(sim.t.141)
df.sim.t.141 <- data.frame(sim.t.141)
write_xlsx(df.sim.t.141,"C:\\Users\\estit\\Documents\\PHD\\
ABERDEEN_1stYear\\R\\sim.t.141.xlsx")

# KS test simulated and actual discovery waiting
# time distribution
ks.test(t,sim.t.141[,1],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  t and sim.t.141[, 1]
```

```
#D = 0.068338, p-value = 0.8981
#alternative hypothesis: two-sided
ks.test(t,sim.t.141[,100],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  t and sim.t.141[, 100]
#D = 0.1959, p-value = 0.009109
#alternative hypothesis: two-sided
ks.test(t,sim.t.141[,500],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  t and sim.t.141[, 500]
#D = 0.043668, p-value = 0.9993
#alternative hypothesis: two-sided
ks.test(t,sim.t.141[,1000],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  t and sim.t.141[, 1000]
#D = 0.04306, p-value = 0.9995
#alternative hypothesis: two-sided
ks.test(t,sim.t.141[,5000],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  t and sim.t.141[, 5000]
#D = 0.072036, p-value = 0.8592
#alternative hypothesis: two-sided
ks.test(t,sim.t.141[,10000],alternative="two.sided")
#       Two-sample Kolmogorov-Smirnov test
#
#data:  t and sim.t.141[, 10000]
#D = 0.13987, p-value = 0.128
```

```
#alternative hypothesis: two-sided

# Simulated discovery dates taken from 1st, 100th, 500th,
# 1,000th, 5000th, and 10,000th replication
library(lubridate)
date0 <- ymd(as.Date('1967-07-09'))
for (i in 1:141){
date0 <- ymd(as.Date('1967-07-09'))
dat.sim.t.141.1 <- date0+cumsum(sim.t.141[,1])
dat.sim.t.141.100 <- date0+cumsum(sim.t.141[,100])
dat.sim.t.141.500 <- date0+cumsum(sim.t.141[,500])
dat.sim.t.141.1000 <- date0+cumsum(sim.t.141[,1000])
dat.sim.t.141.5000 <- date0+cumsum(sim.t.141[,5000])
dat.sim.t.141.10000 <- date0+cumsum(sim.t.141[,10000])
}
df.sim.t.141 <- data.frame(dat.sim.t.141.1,dat.sim.t.141.100,
dat.sim.t.141.500,dat.sim.t.141.1000,dat.sim.t.141.5000,
dat.sim.t.141.10000)
View(df.sim.t.141)
write_xlsx(df.sim.t.141,"C:\\Users\\estit\\Documents\\PHD\\
ABERDEEN_1stYear\\R\\dat.sim.t.141.xlsx")

# Quantile of waiting time
qe <- rowQuantiles(sim.t.141, probs = probs)
View(qe)
df.qe <- data.frame(qe)
write_xlsx(df.qe,"C:\\Users\\estit\\Documents\\PHD\\
ABERDEEN_1stYear\\R\\quantile_exp141.xlsx")

#MAPE test for simulated waiting time
mape(t[t>0],sim.t.141[1:139,1])
```

```
# [1] 3.992272
mape(t[t>0],sim.t.141[1:139,100])
# [1] 3.356068
mape(t[t>0],sim.t.141[1:139,500])
# [1] 4.374441
mape(t[t>0],sim.t.141[1:139,1000])
# [1] 3.084268
mape(t[t>0],sim.t.141[1:139,5000])
# [1] 3.395689
mape(t[t>0],sim.t.141[1:139,10000])
# [1] 5.127476

# Quantile of simulated discovery dates
date0 <- ymd(as.Date('1967-07-09'))
for (i in 1:141){
     date0 <- ymd(as.Date('1967-07-09'))
     dat.qe.10 <- date0+cumsum(qe[,1])
     dat.qe.25 <- date0+cumsum(qe[,2])
     dat.qe.37.5 <- date0+cumsum(qe[,3])
     dat.qe.50 <- date0+cumsum(qe[,4])
     dat.qe.62.5 <- date0+cumsum(qe[,5])
     dat.qe.75 <- date0+cumsum(qe[,6])
     dat.qe.90 <- date0+cumsum(qe[,7])
}
df.qe <- data.frame(dat.qe.10,dat.qe.25,dat.qe.37.5,dat.qe.50,
dat.qe.62.5,dat.qe.75,dat.qe.90)
View(df.qe)
write_xlsx(df.qe,"C:\\Users\\estit\\Documents\\PHD\\
ABERDEEN_1stYear\\R\\quantile_datexp141.xlsx")

# Changing parameters
```

```
# low rate
set.seed(412)
el <- replicate(n = 10000, expr = rexp(n = 141,
rate = 0.0068))
par(mfrow=c(1,1))
hist(el)
df.el <- data.frame(el)
write_xlsx(df.el,"C:\\Users\\estit\\Documents\\PHD\\
ABERDEEN_1stYear\\R\\el.xlsx")

# KS test simulated and actual discovery waiting time
# distribution
ks.test(t,el[,1],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  t and el[, 1]
#D = 0.12401, p-value = 0.2301
#alternative hypothesis: two-sided
ks.test(t,el[,100],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  t and el[, 100]
#D = 0.10274, p-value = 0.4487
#alternative hypothesis: two-sided
ks.test(t,el[,500],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  t and el[, 500]
#D = 0.082877, p-value = 0.7201
#alternative hypothesis: two-sided
ks.test(t,el[,1000],alternative="two.sided")
```

```
#Two-sample Kolmogorov-Smirnov test
#
#data:  t and el[, 1000]
#D = 0.074316, p-value = 0.8326
#alternative hypothesis: two-sided
ks.test(t,el[,5000],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  t and el[, 5000]
#D = 0.081358, p-value = 0.7411
#alternative hypothesis: two-sided
ks.test(t,el[,10000],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  t and el[, 10000]
#D = 0.15344, p-value = 0.07317
#alternative hypothesis: two-sided

# Simulated discovery dates taken from 1st, 100th, 500th,
# 1,000th, 5000th, and 10,000th replication
for (i in 1:141){
date0 <- ymd(as.Date('1967-07-09'))
dat.el.1 <- date0+cumsum(el[,1])
dat.el.100 <- date0+cumsum(el[,100])
dat.el.500 <- date0+cumsum(el[,500])
dat.el.1000 <- date0+cumsum(el[,1000])
dat.el.5000 <- date0+cumsum(el[,5000])
dat.el.10000 <- date0+cumsum(el[,10000])
}
df.el <- data.frame(dat.el.1,dat.el.100,dat.el.500,
dat.el.1000,dat.el.5000,dat.el.10000)
```

```
View(df.el)
write_xlsx(df.el,"C:\\Users\\estit\\Documents\\PHD\\
ABERDEEN_1stYear\\R\\dat.el.xlsx")

# Quantile of waiting time
qel <- rowQuantiles(el, probs = probs)
View(qel)
df.qel <- data.frame(qel)
write_xlsx(df.qel,"C:\\Users\\estit\\Documents\\PHD\\
ABERDEEN_1stYear\\R\\quantile_el.xlsx")

#MAPE test for simulated waiting time
mape(t[t>0],el[1:139,1])
# [1] 3.526484
mape(t[t>0],el[1:139,100])
# [1] 3.745263
mape(t[t>0],el[1:139,500])
# [1] 3.452338
mape(t[t>0],el[1:139,1000])
# [1] 3.031484
mape(t[t>0],el[1:139,5000]])
# [1] 5.208759
mape(t[t>0],el[1:139,10000])
# [1] 5.233531

# Quantile of simulated discovery dates
for (i in 1:141){
     date0 <- ymd(as.Date('1967-07-09'))
     dat.qel.10 <- date0+cumsum(qel[,1])
     dat.qel.25 <- date0+cumsum(qel[,2])
     dat.qel.37.5 <- date0+cumsum(qel[,3])
```

```
      dat.qel.50 <- date0+cumsum(qel[,4])
      dat.qel.62.5 <- date0+cumsum(qel[,5])
      dat.qel.75 <- date0+cumsum(qel[,6])
      dat.qel.90 <- date0+cumsum(qel[,7])
}
df.qel <- data.frame(dat.qel.10,dat.qel.25,dat.qel.37.5,
dat.qel.50,dat.qel.62.5,dat.qel.75,dat.qel.90)
View(df.qel)
write_xlsx(df.qel,"C:\\Users\\estit\\Documents\\PHD\\
ABERDEEN_1stYear\\R\\quantile_datel.xlsx")

# high rate
set.seed(413)
eh <- replicate(n = 10000, expr = rexp(n = 141,
rate = 0.0090))
par(mfrow=c(1,1))
hist(eh)
df.eh <- data.frame(eh)
write_xlsx(df.eh,"C:\\Users\\estit\\Documents\\PHD\\
ABERDEEN_1stYear\\R\\eh.xlsx")

# KS test simulated and actual discovery waiting time
# distribution
ks.test(t,eh[,1],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  t and eh[, 1]
#D = 0.079433, p-value = 0.7672
#alternative hypothesis: two-sided
ks.test(t,eh[,100],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
```

```
#
#data:  t and eh[, 100]
#D = 0.12579, p-value = 0.2163
#alternative hypothesis: two-sided
ks.test(t,eh[,500],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  t and eh[, 500]
#D = 0.082928, p-value = 0.7194
#alternative hypothesis: two-sided
ks.test(t,eh[,1000],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  t and eh[, 1000]
#D = 0.15187, p-value = 0.07827
#alternative hypothesis: two-sided
ks.test(t,eh[,5000],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  t and eh[, 5000]
#D = 0.050557, p-value = 0.9939
#alternative hypothesis: two-sided
ks.test(t,eh[,10000],alternative="two.sided")
#Two-sample Kolmogorov-Smirnov test
#
#data:  t and eh[, 10000]
#D = 0.10046, p-value = 0.4776
#alternative hypothesis: two-sided

# Simulated discovery dates taken from 1st, 100th,
500th, 1,000th, 5000th, and 10,000th replication
```

```
for (i in 1:141){
date0 <- ymd(as.Date('1967-07-09'))
dat.eh.1 <- date0+cumsum(eh[,1])
dat.eh.100 <- date0+cumsum(eh[,100])
dat.eh.500 <- date0+cumsum(eh[,500])
dat.eh.1000 <- date0+cumsum(eh[,1000])
dat.eh.5000 <- date0+cumsum(eh[,5000])
dat.eh.10000 <- date0+cumsum(eh[,10000])
}
df.eh <- data.frame(dat.eh.1,dat.eh.100,dat.eh.500,
dat.eh.1000,dat.eh.5000,dat.eh.10000)
View(df.eh)
write_xlsx(df.eh,"C:\\Users\\estit\\Documents\\PHD\\
ABERDEEN_1stYear\\R\\dat.eh.xlsx")

# Quantile of waiting time
qeh <- rowQuantiles(eh, probs = probs)
View(qeh)
df.qeh <- data.frame(qeh)
write_xlsx(df.qeh,"C:\\Users\\estit\\Documents\\PHD\\
ABERDEEN_1stYear\\R\\quantile_eh.xlsx")

#MAPE test for simulated waiting time
mape(t[t>0],eh[1:139,1])
# [1] 2.760516
mape(t[t>0],eh[1:139,100])
# [1] 3.792903
mape(t[t>0],eh[1:139,500])
# [1] 2.576815
mape(t[t>0],eh[1:139,1000])
# [1] 2.147273
```

```
mape(t[t>0],eh[1:139,5000])
# [1] 4.13765
mape(t[t>0],eh[1:139,10000])
# [1] 3.230472

# Quantile of simulated discovery dates
for (i in 1:141){
     date0 <- ymd(as.Date('1967-07-09'))
     dat.qeh.10 <- date0+cumsum(qeh[,1])
     dat.qeh.25 <- date0+cumsum(qeh[,2])
     dat.qeh.37.5 <- date0+cumsum(qeh[,3])
     dat.qeh.50 <- date0+cumsum(qeh[,4])
     dat.qeh.62.5 <- date0+cumsum(qeh[,5])
     dat.qeh.75 <- date0+cumsum(qeh[,6])
     dat.qeh.90 <- date0+cumsum(qeh[,7])
}
df.qeh <- data.frame(dat.qeh.10,dat.qeh.25,dat.qeh.37.5,
dat.qeh.50,dat.qeh.62.5,dat.qeh.75,dat.qeh.90)
View(df.qeh)
write_xlsx(df.qeh,"C:\\Users\\estit\\Documents\\PHD\\
ABERDEEN_1stYear\\R\\quantile_dateh.xlsx")

# Density for various parameters
par(mfrow=c(1,1))
plot(density(el), main="Density␣simulated␣waiting␣time
(Exponential␣distribution)",col="blue",xlim=c(0,1000),
ylim=c(0,0.009))
lines(density(sim.t.141),col = "red")
lines(density(eh),col = "orange")
legend("topright", c("k=0.0068", "k=0.0073", "k=0.0090"),
col = c("blue","red","orange"), lty = 1)
```

```
# Plot simulated discovery size and waiting time taken from
# 1st, 100th, 500th, 1,000th, 5000th, and 10,000th
replication
# NCS parameter
par(mfrow=c(3,2))
plot(dat.sim.t.141.1,sim.ln.141[,1],xlim=as.Date(c
("1967-07-09","2030-12-31")),ylim=c(0,1800),type="line",
col="dark␣blue",xlab="Simulated␣discovery␣year
(1st␣replication)",ylab="Simulated␣discovery␣size␣(mill
Sm3)")
plot(dat.sim.t.141.100,sim.ln.141[,100],xlim=as.Date(c
("1967-07-09","2030-12-31")),ylim=c(0,1800),type="line",
col="purple",xlab="Simulated␣discovery␣year␣(100th
replication)",ylab="Simulated␣discovery␣size␣(mill␣Sm3)")
plot(dat.sim.t.141.500,sim.ln.141[,500],xlim=as.Date(c
("1967-07-09","2030-12-31")),ylim=c(0,1800),type="line",
col="dark␣green",xlab="Simulated␣discovery␣year␣(500th
replication)",ylab="Simulated␣discovery␣size␣(mill␣Sm3)")
plot(dat.sim.t.141.1000,sim.ln.141[,1000],xlim=as.Date(c
("1967-07-09",
"2030-12-31")),ylim=c(0,1800),type="line",col="orange",
xlab="Simulated␣discovery␣year␣(1,000th␣replication)",
ylab="Simulated␣discovery␣size␣(mill␣Sm3)")
plot(dat.sim.t.141.5000,sim.ln.141[,5000],xlim=as.Date(c
("1967-07-09",
"2030-12-31")),ylim=c(0,1800),type="line",col="pink",
xlab="Simulated␣discovery␣year␣(5,000th␣replication)",
ylab="Simulated␣discovery␣size␣(mill␣Sm3)")
plot(dat.sim.t.141.10000,sim.ln.141[,10000],xlim=as.Date(c
("1967-07-09","2030-12-31")),ylim=c(0,1800),type="line",
```

```
col="dark␣red",xlab="Simulated␣discovery␣year␣(10,000th
replication)",ylab="Simulated␣discovery␣size␣(mill␣Sm3)")

# low parameter (discovery size ln.l; waiting time el)
par(mfrow=c(3,2))
plot(dat.el.1,ln.l[,1],xlim=as.Date(c("1967-07-09",
"2030-12-31")),ylim=c(0,500),type="line",col="dark␣blue",
xlab="Simulated␣discovery␣year␣(1st␣replication)",
ylab="Simulated␣discovery␣size␣(mill␣Sm3)")
plot(dat.el.100,ln.l[,100],xlim=as.Date(c("1967-07-09",
"2030-12-31")),ylim=c(0,500),type="line",col="purple",
xlab="Simulated␣discovery␣year␣(100th␣replication)",
ylab="Simulated␣discovery␣size␣(mill␣Sm3)")
plot(dat.el.500,ln.l[,500],xlim=as.Date(c("1967-07-09",
"2030-12-31")), ylim=c(0,500),type="line",col="dark␣green",
xlab="Simulated␣discovery␣year␣(500th␣replication)",
ylab="Simulated␣discovery␣size␣(mill␣Sm3)")
plot(dat.el.1000,ln.l[,1000],xlim=as.Date(c("1967-07-09",
"2030-12-31")),ylim=c(0,500),type="line",col="orange",
xlab="Simulated␣discovery␣year␣(1000th␣replication)",
ylab="Simulated␣discovery␣size␣(mill␣Sm3)")
plot(dat.el.5000,ln.l[,5000],xlim=as.Date(c("1967-07-09",
"2030-12-31")),ylim=c(0,500),type="line",col="pink",
xlab="Simulated␣discovery␣year␣(5,000th␣replication)",
ylab="Simulated␣discovery␣size␣(mill␣Sm3)")
plot(dat.el.10000,ln.l[,10000],xlim=as.Date(c("1967-07-09",
"2030-12-31")),ylim=c(0,500),type="line",col="dark␣red",
xlab="Simulated␣discovery␣year␣(10,000th␣replication)",
ylab="Simulated␣discovery␣size␣(mill␣Sm3)")

# high parameter (discovery size ln.h; waiting time eh)
```

```
par(mfrow=c(3,2))
plot(dat.eh.1,ln.h[,1],xlim=as.Date(c("1967-07-09",
"2030-12-31")), ylim=c(0,3000),type="line",col=
"dark␣blue",xlab="Simulated␣discovery␣year
(1st␣replication)",ylab="Simulated␣discovery␣size
(mill␣Sm3)")
plot(dat.eh.100,ln.h[,100],xlim=as.Date(c("1967-07-09",
"2030-12-31")), ylim=c(0,3000),type="line",col="purple",
xlab="Simulated␣discovery␣year␣(100th␣replication)",
ylab="Simulated␣discovery␣size␣(mill␣Sm3)")
plot(dat.eh.500,ln.h[,500],xlim=as.Date(c("1967-07-09",
"2030-12-31")), ylim=c(0,3000),type="line",col=
"dark␣green", xlab="Simulated␣discovery␣year
(500th␣replication)", ylab="Simulated␣discovery
size␣(mill␣Sm3)")
plot(dat.eh.1000,ln.h[,1000],xlim=as.Date(c("1967-07-09",
"2030-12-31")),ylim=c(0,3000),type="line",col="orange",
xlab="Simulated␣discovery␣year␣(1000th␣replication)",
ylab="Simulated␣discovery␣size␣(mill␣Sm3)")
plot(dat.eh.5000,ln.h[,5000],xlim=as.Date(c("1967-07-09",
"2030-12-31")),ylim=c(0,3000),type="line",col="pink",
xlab="Simulated␣discovery␣year␣(5,000th␣replication)",
ylab="Simulated␣discovery␣size␣(mill␣Sm3)")
plot(dat.eh.10000,ln.h[,10000],xlim=as.Date
(c("1967-07-09","2030-12-31")),ylim=c(0,3000),
type="line",col="dark␣red",xlab="Simulated␣discovery
year␣(10,000th␣replication)",ylab="Simulated␣discovery
size␣(mill␣Sm3)")
\end{singlespace}
```

A.2 Robustness test

Table A.1 reports the ARDL with a structural break in the error correction form by adjusting the nominal oil price with the PPP of Norway to obtain real oil price. Column (i) presents the ARDL estimates in error correction form of waiting time between discoveries as the dependent variable, and Column (ii) shows the estimates of discovery size as a dependent variable. Panel (i) reports that the F-statistic and t-statistic of the adjustment factors are larger than the bounds test critical values at a 1% and 10% significance level for waiting time and discovery size, respectively. The null hypothesis that the level relationship does not exist is rejected; hence there is enough evidence to support the long-run relationship at the level of waiting time, exploration well counts, and real oil price, and so does for discovery size, exploration well, and real oil price.

The adjustment factor ($-\alpha$) shows how the dependent variable changes when the three variables deviate from their long-run equilibrium. The adjustment factors for the two equations are all negative and statistically significant, which means that the estimated error correction forms represent stable relationships. The dependent variable of the prior period is too high relative to the long-run equilibrium, so it is necessary to decrease its value in the current period to revert to equilibrium. The long-run equilibrium relationship denoted by coefficient θ in Table A.1 represents a contemporaneous relationship between variables. The sign of the coefficient estimates is consistent with the model applying U.S. CPI as an adjustment inflation rate of oil price. A 1% real oil price relates negatively with the waiting time by 0.2 days and positively with discovery size by 0.31%. A 1% increase in exploration well counts relates negatively with the waiting time by 0.9 days and with discovery size by 0.1%.

Table A.1: ARDL estimation with a structural break in error correction form for exploratory effort, efficiency, and adjusted PPP real oil price

	(i) D.wt$_t$	(ii) D.ls$_t$
(i) *Case 3*		
Bound Test H$_0$: no level relationship		
F-stat‡	6.024*	47.203***
t-stat‡	-3.382*	-11.057***
(ii) *Adjustment factor* *(-α)*‡		
wt$_{t-1}$	-0.666*** (0.197)	
ls$_{t-1}$		-0.812*** (0.074)
(iii) *Long-run* (θ)		
lw$_{t-1}$	-93.123 (80.750)	-0.113 (0.286)
lop$_{t-1}$	-20.790 (51.338)	0.310* (0.172)
(iv) *short-run* ($\psi_{yi}, \omega^{'}, \psi^{'}_{xi}$)		
D.wt$_{t-1}$	-0.276** (0.126)	
D.ls$_{t-1}$		
D.ls$_{t-2}$		
D.ls$_{t-3}$		
D.lw$_t$	-64.311 (62.221)	-0.092 (0.232)
D.lw$_{t-1}$	-138.679** (64.022)	
D.lw$_{t-2}$	-143.843*** (50.575)	
D.lw$_{t-3}$		
D.lop$_t$	-13.849 (36.184)	1.012*** (0.223)
b$_t$	9.218 (51.960)	-0.582*** (0.187)
(a_0) Constant	387.750* (224.679)	-0.075 (0.602)
Observations	50	50

Notes:
Standard error in parentheses; * $p < .10$, ** $p < .05$, *** $p < .01$
‡The approximate p-values applied for speed-of-adjustment coefficient is based on Kripfganz et al. (2018)

Appendix B

Global oil market uncertainty, oil exploration, and crude oil price: An application of Google Trends

B.1 The optimal lag length selection

The optimal lag length selection is chosen based on information criteria. Apart from information criteria explained earlier in Equation 2.26 and 2.27, the sequential modified likelihood ratio (LR) test can be applied in the VAR estimation. The LR test is applied based on the log likelihood function and has the hypothesis that lag l coefficients are jointly zero using Wald (χ^2) statistic (Lütkepohl, 1991, cited by EViews, 2019). The test starts from the maximum lag then decreases by one lag until the hypothesis can be rejected at 5% critical value (which is marked by an asterisk in Table B.1).

The optimal lag chosen is the smallest value of information criteria. First, AIC and SC are considered, however if the optimal lag obtained by the AIC or SC is too short (e.g. 1 or 2 lags), the lag obtained by the LR test is applied for the VAR estimation to avoid autocorrelation of the model.

The endogenous variables applied in VAR estimates are in log-differences. Figure B.1 and B.2 show the time series of uncertainty, rig counts, and crude oil price in log-differences.

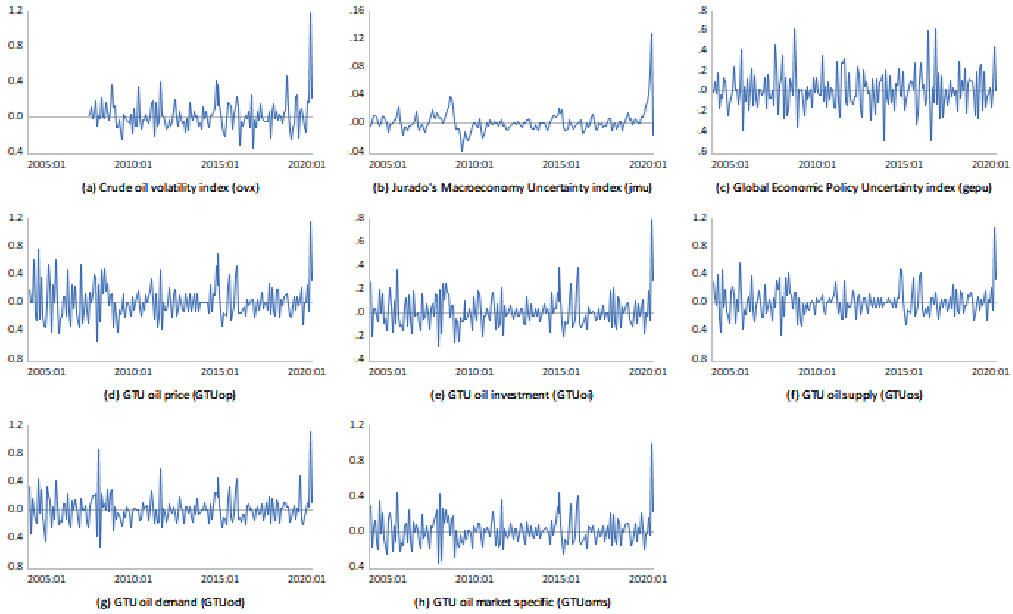
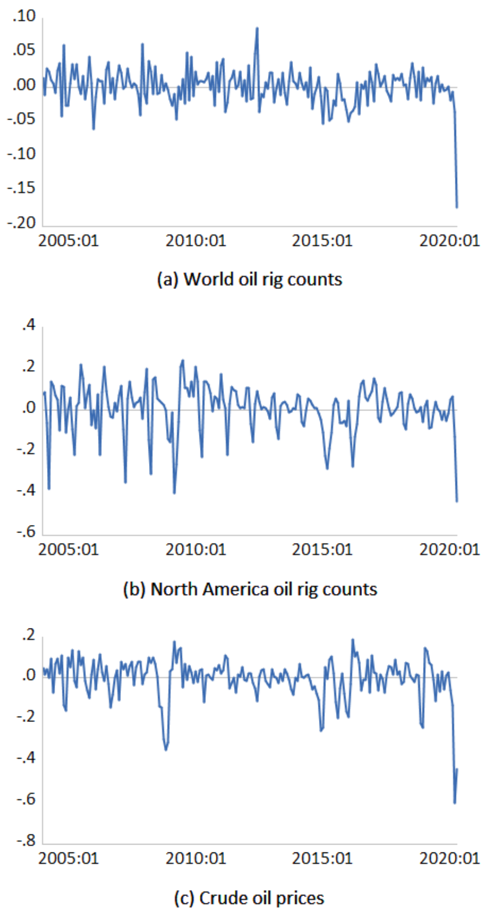


Figure B.1: Uncertainty indices (log differences)

The equation of the LR test is given by,

$$LR = (T-m)\left\{log\left|\sum \epsilon, l-1\right| - log\left|\sum \epsilon, l\right|\right\} \chi^2(k^2) \tag{B.1}$$

where T is the number of observation, p is the number of the lags, k is the number of the endogenous variables, m is the number of parameters estimated in each equation and $m = (pk + d)$ (refer to Equation 2.24); Σ_ϵ is the determinant of the residual covariance, l is the log of the likelihood function.



(a) World oil rig counts

(b) North America oil rig counts

(c) Crude oil prices

Figure B.2: Oil market variables (log differences)

Table B.1: VAR optimal lag length based on information criteria

(a) World exploration model

VAR equation	Lag	LR	AIC	SC
rig_w op jmu	6	29.815*	-13.994*	-13.318
rig_w op ovx	6	23.463*	-7.529*	-7.595*
rig_w op gepu	6	27.877*	-7.114*	-6.407
rig_w op gtu_{op}	6	31.285*	-6.219*	-5.325
rig_w op gtu_{oi}	7	24.962*	-7.348*	-6.310
rig_w op gtu_{os}	8	19.373*	-6.643*	-5.460
rig_w op gtu_{od}	6	30.028*	-6.795*	-5.901
rig_w op gtu_{oms}	7	21.626*	-7.045*	-6.007

(b) North America exploration model

VAR equation	Lag	LR	AIC	SC
rig_w op jmu	12	54.618*	-11.866*	-10.761
rig_w op ovx	12	40.383*	-5.736*	-3.680
rig_w op gepu	12	35.263*	-4.452*	-3.055
rig_w op gtu_{op}	12	62.562*	-4.154*	-2.379
rig_w op gtu_{oi}	12	64.300*	-5.334*	-3.559
rig_w op gtu_{os}	12	63.941*	-4.552*	-2.777
rig_w op gtu_{od}	12	59.046*	-4.764*	-2.989
rig_w op gtu_{oms}	12	64.613*	-5.008*	-3.233

B.2 Unit root test

Table B.2 presents the unit root test based on Narayan and Popp (2010). The result indicates stationary results for all variables, including oil price in level, which is a sceptical result, as oil prices have been through many fluctuation periods.

Table B.2: Narayan and Popp (2010) Unit Root Test for Uncertainty and Oil Market Variables

	Level	Log-level	First difference	Log-first difference
JMU	-5.653***	-5.662***	-12.329***	-11.958***
OVX	-6.596***	-6.636***	-13.917***	-12.711***
GEPU	-7.859***	-7.488***	-13.223***	-11.001***
GTU_{op}	-7.129***	-6.725***	-18.157***	-17.197***
GTU_{oi}	-7.847***	-6.967***	-17.142***	-14.414***
GTU_{od}	-6.669***	-6.431***	-20.316***	-14.064***
GTU_{os}	-7.451***	-6.781***	-18.204***	-15.148***
GTU_{oms}	-7.065***	-6.654***	-19.113***	-15.867***
rig_{w}	-6.588***	-6.737***	-17.898***	-17.806***
rig_{na}	-6.728***	-5.795***	-10.896***	-8.476***
op	-5.918***	-5.192**	-15.799***	-17.190***

B.3 VAR estimates

Tables B.3 - B.4 show the significance of the dynamic lag of the independent variables in VAR estimates.

Table B.3: VAR estimates for world exploration activity

(a) Endogenous variables: rig_w op ovx

	$\text{rig}_{w,t}$	op_t	ovx_t
$\text{rig}_{w,t-1}$	-0.026 (0.072)	-0.316 (0.275)	0.702 (0.505)
$\text{rig}_{w,t-2}$	0.037 (0.067)	0.109 (0.258)	-0.499 (0.474)
$\text{rig}_{w,t-3}$	0.072 (0.066)	0.253 (0.254)	0.859* (0.467)
$\text{rig}_{w,t-4}$	0.077 (0.069)	-0.099 (0.264)	0.301 (0.485)
$\text{rig}_{w,t-5}$	0.064 (0.069)	0.026 (0.266)	-0.534 (0.489)
$\text{rig}_{w,t-6}$	0.191*** (0.069)	-0.501* (0.266)	0.860* (0.488)
op_{t-1}	0.024 (0.026)	0.431*** (0.100)	-0.188 (0.184)
op_{t-2}	0.008 (0.028)	-0.012 (0.107)	-0.040 (0.197)
op_{t-3}	0.058** (0.028)	0.059 (0.107)	-0.215 (0.197)
op_{t-4}	-0.036 (0.029)	-0.049 (0.111)	0.257 (0.204)
op_{t-5}	0.073** (0.028)	-0.125 (0.108)	0.161 (0.199)
op_{t-6}	0.062** (0.025)	-0.063 (0.095)	-0.051 (0.174)
ovx_{t-1}	-0.037*** (0.014)	-0.132** (0.055)	0.061 (0.101)
ovx_{t-2}	-0.044*** (0.015)	0.131** (0.057)	-0.161 (0.104)
ovx_{t-3}	-0.013* (0.015)	0.131** (0.057)	-0.220** (0.106)
ovx_{t-4}	-0.010* (0.015)	0.012 (0.056)	-0.090 (0.103)
ovx_{t-5}	0.005 (0.015)	-0.065 (0.056)	0.079 (0.104)
ovx_{t-6}	-0.008 (0.014)	-0.066 (0.055)	0.021 (0.101)
Constant	6.8e-5** (0.002)	-0.001 (0.007)	0.005 (0.012)
Observations	175	175	175
RMSE	0.024	0.093	0.170

Notes:
Standard errors are in parentheses * $p < .10$, ** $p < .05$, *** $p < .01$

(b) Endogenous variables: rig_w op jmu

	$\text{rig}_{w,t}$	op_t	jmu_t
$\text{rig}_{w,t-1}$	-0.041 (0.054)	-0.141 (0.173)	0.005 (0.013)
$\text{rig}_{w,t-2}$	0.025 (0.053)	-0.192 (0.169)	0.001 (0.012)
$\text{rig}_{w,t-3}$	0.116** (0.052)	0.191 (0.167)	-3.5e-4 (0.012)
$\text{rig}_{w,t-4}$	0.064 (0.052)	-0.032 (0.166)	0.012 (0.012)
$\text{rig}_{w,t-5}$	0.092* (0.052)	-0.056 (0.167)	-0.001 (0.012)
$\text{rig}_{w,t-6}$	0.167*** (0.052)	-0.231 (0.165)	0.018 (0.012)
op_{t-1}	0.046** (0.018)	0.410*** (0.056)	0.010** (0.004)
op_{t-2}	0.025 (0.020)	-0.164*** (0.062)	0.005 (0.005)
op_{t-3}	0.058*** (0.019)	-0.002 (0.062)	0.003 (0.005)
op_{t-4}	-0.008 (0.020)	-0.062 (0.062)	0.007 (0.005)
op_{t-5}	0.030 (0.019)	0.021 (0.060)	-0.002 (0.004)
op_{t-6}	0.048*** (0.017)	-0.034 (0.055)	0.006 (0.004)
jmu_{t-1}	-0.366 (0.239)	-4.921*** (0.761)	0.861*** (0.056)
jmu_{t-2}	-0.301 (0.327)	1.572 (1.043)	-0.025 (0.077)
jmu_{t-3}	-0.069 (0.327)	2.444** (1.042)	-0.058 (0.077)
jmu_{t-4}	-0.084 (0.329)	-2.208** (1.050)	-0.037 (0.077)
jmu_{t-5}	0.524 (0.332)	0.854 (1.058)	0.030 (0.078)
jmu_{t-6}	-0.435* (0.262)	-1.210 (0.834)	0.002 (0.061)
Constant	3.1e-4 (0.001)	0.004 (0.004)	4.5e-5 (3.3e-4)
Observations	317	317	317
RMSE	0.025	0.080	0.006

Notes:
Standard errors are in parentheses * $p < .10$, ** $p < .05$, *** $p < .01$

(c) Endogenous variables: rig_w op gepu

	$\text{rig}_{w,t}$	op_t	gepu_t
$\text{rig}_{w,t-1}$	0.003 (0.057)	-0.147 (0.201)	-0.544 (0.371)
$\text{rig}_{w,t-2}$	0.040 (0.056)	-0.097 (0.199)	-1.098*** (0.367)
$\text{rig}_{w,t-3}$	0.139** (0.056)	0.154 (0.199)	-0.273 (0.367)
$\text{rig}_{w,t-4}$	0.052 (0.055)	-0.068 (0.196)	0.126 (0.361)
$\text{rig}_{w,t-5}$	0.067 (0.055)	-0.064 (0.195)	0.803** (0.360)
$\text{rig}_{w,t-6}$	0.155*** (0.054)	-0.257 (0.192)	1.125*** (0.355)
op_{t-1}	0.070*** (0.016)	0.466*** (0.059)	-0.158 (0.109)
op_{t-2}	0.048*** (0.019)	-0.206*** (0.066)	0.024 (0.122)
op_{t-3}	0.054*** (0.019)	-0.028 (0.068)	0.307** (0.125)
op_{t-4}	-0.020 (0.019)	-0.049 (0.069)	0.052 (0.128)
op_{t-5}	0.037** (0.019)	0.008 (0.067)	0.293** (0.124)
op_{t-6}	0.043** (0.017)	0.003 (0.062)	-0.081 (0.114)
gepu_{t-1}	0.002 (0.009)	-0.054* (0.031)	-0.270*** (0.058)
gepu_{t-2}	0.012 (0.009)	-0.033 (0.032)	-0.240*** (0.059)
gepu_{t-3}	0.025*** (0.009)	-0.012 (0.034)	-0.162*** (0.062)
gepu_{t-4}	0.007 (0.009)	-0.038 (0.033)	-0.036 (0.061)
gepu_{t-5}	0.002 (0.009)	-0.038 (0.033)	-0.035 (0.060)
gepu_{t-6}	0.009 (0.009)	0.033 (0.032)	0.050 (0.058)
Constant	-9.3e-4 (0.001)	0.004 (0.005)	0.007 (0.009)
Observations	298	298	298
RMSE	0.025	0.090	0.165

Notes:
Standard errors are in parentheses * $p < .10$, ** $p < .05$, *** $p < .01$

(d) Endogenous variables: rig_w op gtu_{op}

	$\text{rig}_{w,t}$	op_t	$\text{gtu}_{op,t}$
$\text{rig}_{w,t-1}$	-0.022 (0.065)	-0.343 (0.239)	0.737 (0.663)
$\text{rig}_{w,t-2}$	0.022 (0.064)	0.031 (0.234)	-1.188* (0.650)
$\text{rig}_{w,t-3}$	0.071 (0.064)	0.008 (0.233)	0.717 (0.648)
$\text{rig}_{w,t-4}$	0.048 (0.064)	0.140 (0.232)	0.525 (0.646)
$\text{rig}_{w,t-5}$	0.041 (0.064)	0.072 (0.233)	0.645 (0.647)
$\text{rig}_{w,t-6}$	0.208*** (0.063)	-0.244 (0.229)	-0.047 (0.636)
op_{t-1}	0.054*** (0.019)	0.528*** (0.068)	-0.104 (0.190)
op_{t-2}	0.042** (0.021)	-0.236*** (0.078)	0.299 (0.216)
op_{t-3}	0.061*** (0.022)	-0.023 (0.079)	0.098 (0.219)
op_{t-4}	-0.016 (0.022)	-7.7e-4 (0.080)	0.114 (0.222)
op_{t-5}	0.037* (0.022)	-0.069 (0.079)	0.292 (0.218)
op_{t-6}	0.063*** (0.019)	-0.022 (0.070)	0.062 (0.193)
$\text{gtu}_{op,t-1}$	-0.018*** (0.007)	-0.075*** (0.025)	-0.108 (0.069)
$\text{gtu}_{op,t-2}$	-0.014** (0.007)	0.024 (0.026)	-0.077 (0.072)
$\text{gtu}_{op,t-3}$	-4e-4 (0.007)	-0.024 (0.026)	-0.020 (0.071)
$\text{gtu}_{op,t-4}$	0.001 (0.007)	-0.018 (0.026)	-0.082 (0.071)
$\text{gtu}_{op,t-5}$	-0.009 (0.007)	0.015 (0.026)	0.011 (0.072)
$\text{gtu}_{op,t-6}$	-0.005 (0.007)	-0.074*** (0.025)	0.096 (0.071)
Constant	-3e-4 (0.002)	0.004 (0.006)	0.009 (0.017)
Observations	215	215	215
RMSE	0.025	0.091	0.253

Notes:
Standard errors are in parentheses * $p < .10$, ** $p < .05$, *** $p < .01$

(e) Endogenous variables: rig_w op gtu_{oi}

	$\text{rig}_{w,t}$	op_t	$\text{gtu}_{oi,t}$
$\text{rig}_{w,t-1}$	-0.046 (0.067)	-0.494** (0.243)	0.370 (0.407)
$\text{rig}_{w,t-2}$	-0.003 (0.065)	0.017 (0.231)	-0.571 (0.389)
$\text{rig}_{w,t-3}$	0.066 (0.064)	-0.058 (0.226)	0.529 (0.380)
$\text{rig}_{w,t-4}$	0.057 (0.064)	0.127 (0.227)	0.369 (0.382)
$\text{rig}_{w,t-5}$	0.038 (0.064)	0.020 (0.228)	0.340 (0.383)
$\text{rig}_{w,t-6}$	0.203*** (0.065)	-0.403* (0.228)	0.073 (0.383)
$\text{rig}_{w,t-7}$	0.058 (0.065)	0.513** (0.231)	-0.199 (0.388)
op_{t-1}	0.057*** (0.019)	0.493*** (0.067)	-0.152 (0.112)
op_{t-2}	0.037* (0.021)	-0.204*** (0.076)	0.271** (0.127)
op_{t-3}	0.062*** (0.022)	-0.034 (0.077)	-0.091 (0.130)
op_{t-4}	-0.006 (0.022)	-0.012 (0.078)	0.223* (0.131)
op_{t-5}	0.040* (0.022)	-0.009 (0.077)	0.062 (0.130)
op_{t-6}	0.063*** (0.022)	-0.118 (0.077)	0.007 (0.129)
op_{t-7}	0.015 (0.020)	0.053 (0.069)	0.051 (0.116)
$\text{gtu}_{oi,t-1}$	-0.035*** (0.012)	-0.138*** (0.041)	-0.087 (0.069)
$\text{gtu}_{oi,t-2}$	-0.027** (0.012)	0.050 (0.043)	-0.193*** (0.073)
$\text{gtu}_{oi,t-3}$	-0.016 (0.013)	-0.069 (0.046)	-0.038 (0.077)
$\text{gtu}_{oi,t-4}$	-0.005 (0.013)	-0.047 (0.046)	-0.148* (0.078)
$\text{gtu}_{oi,t-5}$	-0.005 (0.013)	4.9e-5 (0.046)	0.042 (0.077)
$\text{gtu}_{oi,t-6}$	-0.019 (0.013)	-0.156*** (0.046)	0.007 (0.077)
$\text{gtu}_{oi,t-7}$	0.009 (0.013)	-0.170*** (0.045)	-0.090 (0.076)
Constant	-4.6e-4 (0.002)	0.004 (0.006)	0.003 (0.010)
Observations	214	214	214
RMSE	0.025	0.088	0.147

Notes:
Standard errors are in parentheses * $p < .10$, ** $p < .05$, *** $p < .01$

(f) Endogenous variables: rig_w op gtu_{os}

	$\text{rig}_{w,t}$	op_t	$\text{gtu}_{os,t}$
$\text{rig}_{w,t-1}$	-0.046 (0.067)	-0.464* (0.242)	0.359 (0.571)
$\text{rig}_{w,t-2}$	-0.037 (0.068)	0.017 (0.245)	-1.211** (0.577)
$\text{rig}_{w,t-3}$	0.018 (0.065)	0.095 (0.234)	0.735 (0.551)
$\text{rig}_{w,t-4}$	0.056 (0.064)	0.221 (0.231)	0.651 (0.543)
$\text{rig}_{w,t-5}$	0.022 (0.064)	-0.021 (0.231)	0.197 (0.544)
$\text{rig}_{w,t-6}$	0.180*** (0.063)	-0.351 (0.228)	0.276 (0.537)
$\text{rig}_{w,t-7}$	0.003 (0.064)	0.588** (0.232)	-0.688 (0.547)
$\text{rig}_{w,t-8}$	0.047 (0.064)	-0.465** (0.231)	0.012 (0.545)
op_{t-1}	0.044** (0.019)	0.522*** (0.070)	-0.135 (0.165)
op_{t-2}	0.048** (0.021)	-0.200*** (0.076)	0.338* (0.180)
op_{t-3}	0.063*** (0.022)	-0.037 (0.078)	-0.050 (0.183)
op_{t-4}	-0.013 (0.022)	-0.005 (0.080)	0.216 (0.188)
op_{t-5}	0.045** (0.022)	-0.018 (0.079)	0.217 (0.185)
op_{t-6}	0.072*** (0.022)	-0.134* (0.079)	0.008 (0.186)
op_{t-7}	-0.009 (0.022)	0.059 (0.079)	0.061 (0.187)
op_{t-8}	0.063*** (0.019)	-0.008 (0.070)	0.217 (0.165)
$\text{gtu}_{os,t-1}$	-0.029*** (0.008)	-0.094*** (0.030)	-0.022 (0.071)
$\text{gtu}_{os,t-2}$	-0.018** (0.009)	0.036 (0.031)	-0.162** (0.073)
$\text{gtu}_{os,t-3}$	-0.006 (0.009)	-0.048 (0.032)	-0.110 (0.076)
$\text{gtu}_{os,t-4}$	-0.007 (0.009)	-0.012 (0.033)	-0.083 (0.078)
$\text{gtu}_{os,t-5}$	-0.015* (0.009)	0.035 (0.032)	1e-4 (0.076)
$\text{gtu}_{os,t-6}$	-0.009 (0.009)	-0.099*** (0.032)	0.064 (0.075)
$\text{gtu}_{os,t-7}$	-0.007 (0.009)	-0.089*** (0.033)	-0.074 (0.076)
$\text{gtu}_{os,t-8}$	0.003 (0.009)	0.044 (0.032)	0.032 (0.076)
Constant	-6.4e-4 (0.002)	0.003 (0.006)	0.004 (0.014)
Observations	213	213	213
RMSE	0.025	0.089	0.209

Notes:
Standard errors are in parentheses * $p < .10$, ** $p < .05$, *** $p < .01$

(g) Endogenous variables: rig_w op gtu_{od}

	$\text{rig}_{w,t}$	op_t	$\text{gtu}_{od,t}$
$\text{rig}_{w,t-1}$	-0.021 (0.065)	-0.309 (0.242)	0.695 (0.516)
$\text{rig}_{w,t-2}$	0.045 (0.064)	0.034 (0.238)	-0.905 (0.507)
$\text{rig}_{w,t-3}$	0.084 (0.064)	0.009 (0.237)	0.356 (0.505)
$\text{rig}_{w,t-4}$	0.042 (0.063)	0.187 (0.236)	0.348 (0.503)
$\text{rig}_{w,t-5}$	0.039 (0.064)	0.083 (0.236)	0.584 (0.504)
$\text{rig}_{w,t-6}$	0.205*** (0.062)	-0.277 (0.231)	0.380 (0.494)
op_{t-1}	0.047** (0.019)	0.491*** (0.072)	-0.168 (0.152)
op_{t-2}	0.047** (0.021)	-0.201** (0.079)	0.092 (0.169)
op_{t-3}	0.058*** (0.022)	-0.020 (0.080)	0.026 (0.170)
op_{t-4}	-0.023 (0.022)	0.005 (0.082)	0.122 (0.175)
op_{t-5}	0.030 (0.022)	-0.032 (0.081)	0.250 (0.173)
op_{t-6}	0.060*** (0.019)	-0.036 (0.070)	0.180 (0.149)
$\text{gtu}_{od,t-1}$	-0.028*** (0.009)	-0.111*** (0.033)	-0.134* (0.071)
$\text{gtu}_{od,t-2}$	-0.018* (0.010)	0.028 (0.035)	-0.139* (0.075)
$\text{gtu}_{od,t-3}$	0.004 (0.009)	-0.006 (0.035)	-0.097 (0.075)
$\text{gtu}_{od,t-4}$	-9.8e-4 (0.010)	0.002 (0.036)	-0.140* (0.076)
$\text{gtu}_{od,t-5}$	-0.015 (0.010)	0.048 (0.037)	0.070 (0.079)
$\text{gtu}_{od,t-6}$	-0.012 (0.010)	-0.039 (0.036)	0.184** (0.077)
Constant	-2.6e-4 (0.002)	0.003 (0.006)	0.009 (0.013)
Observations	215	215	215
RMSE	0.025	0.091	0.195

Notes:
Standard errors are in parentheses * $p < .10$, ** $p < .05$, *** $p < .01$

(h) Endogenous variables: rig_w op gtu_{oms}

	$\text{rig}_{w,t}$	op_t	$\text{gtu}_{oms,t}$
$\text{rig}_{w,t-1}$	-0.041 (0.072)	-0.488* (0.258)	0.538 (0.505)
$\text{rig}_{w,t-2}$	0.019 (0.069)	-0.054 (0.246)	-0.856* (0.483)
$\text{rig}_{w,t-3}$	0.071 (0.068)	-0.013 (0.242)	0.598 (0.474)
$\text{rig}_{w,t-4}$	0.040 (0.068)	0.215 (0.243)	0.458 (0.476)
$\text{rig}_{w,t-5}$	0.042 (0.068)	-0.011 (0.243)	0.346 (0.477)
$\text{rig}_{w,t-6}$	0.207*** (0.068)	-0.429* (0.242)	0.229 (0.475)
$\text{rig}_{w,t-7}$	0.040 (0.068)	0.509** (0.244)	-0.311 (0.479)
op_{t-1}	0.049** (0.020)	0.489*** (0.072)	-0.143 (0.141)
op_{t-2}	0.041* (0.023)	-0.180** (0.081)	0.214 (0.159)
op_{t-3}	0.062*** (0.023)	-0.046 (0.082)	-0.042 (0.161)
op_{t-4}	-0.015 (0.023)	0.010 (0.083)	0.185 (0.163)
op_{t-5}	0.035 (0.023)	-0.018 (0.083)	0.160 (0.162)
op_{t-6}	0.063*** (0.023)	-0.130 (0.082)	0.043 (0.162)
op_{t-7}	0.015 (0.021)	0.045 (0.074)	0.080 (0.144)
$\text{gtu}_{oms,t-1}$	-0.034*** (0.011)	-0.119*** (0.038)	-0.073 (0.075)
$\text{gtu}_{oms,t-2}$	-0.023** (0.011)	0.041 (0.040)	-0.158** (0.079)
$\text{gtu}_{oms,t-3}$	-0.004 (0.012)	-0.056 (0.041)	-0.089 (0.081)
$\text{gtu}_{oms,t-4}$	-0.004 (0.012)	-0.031 (0.042)	-0.114 (0.082)
$\text{gtu}_{oms,t-5}$	-0.012 (0.012)	0.033 (0.042)	0.061 (0.082)
$\text{gtu}_{oms,t-6}$	-0.014 (0.012)	-0.116*** (0.042)	0.083 (0.083)
$\text{gtu}_{oms,t-7}$	0.007 (0.012)	-0.142*** (0.041)	-0.033 (0.081)
Constant	-3.9e-4 (0.002)	0.004 (0.006)	0.004 (0.012)
Observations	214	214	214
RMSE	0.025	0.088	0.173

Notes:
Standard errors are in parentheses * $p < .10$, ** $p < .05$, *** $p < .01$

Table B.4: VAR estimates for North America exploration activity

(a) Endogenous variables: rig_{na} op ovx

	$\text{rig}_{na,t}$	op_t	ovx
$\text{rig}_{na,t-1}$	0.206*** (0.067)	0.054 (0.121)	0.007 (0.225)
$\text{rig}_{na,t-2}$	-0.135** (0.061)	0.063 (0.110)	-0.113 (0.205)
$\text{rig}_{na,t-3}$	-0.101* (0.061)	-0.084 (0.110)	-0.039 (0.205)
$\text{rig}_{na,t-4}$	-0.031 (0.062)	0.201* (0.111)	-0.047 (0.207)
$\text{rig}_{na,t-5}$	-0.022 (0.062)	-0.147 (0.111)	0.179 (0.207)
$\text{rig}_{na,t-6}$	-0.004 (0.062)	0.185* (0.111)	-0.606*** (0.206)
$\text{rig}_{na,t-7}$	-0.105* (0.062)	-0.062 (0.111)	0.240 (0.206)
$\text{rig}_{na,t-8}$	-0.070 (0.061)	0.238** (0.109)	-0.127 (0.203)
$\text{rig}_{na,t-9}$	-0.037 (0.061)	-0.080 (0.109)	0.116 (0.204)
$\text{rig}_{na,t-10}$	-0.288*** (0.059)	-0.046 (0.107)	0.029 (0.198)
$\text{rig}_{na,t-11}$	0.159** (0.063)	0.114 (0.113)	-0.271 (0.211)
$\text{rig}_{na,t-12}$	0.329*** (0.054)	-0.136 (0.098)	0.233 (0.182)
op_{t-1}	0.137** (0.059)	0.447*** (0.106)	-0.254 (0.197)
op_{t-2}	0.082 (0.064)	-0.020 (0.114)	0.066 (0.213)
op_{t-3}	0.200*** (0.062)	0.003 (0.112)	-0.130 (0.208)
op_{t-4}	0.134 (0.066)	-0.089 (0.118)	0.199 (0.220)
op_{t-5}	0.220*** (0.066)	-0.151 (0.119)	0.285 (0.222)
op_{t-6}	0.119* (0.068)	-0.104 (0.122)	-0.076 (0.227)
op_{t-7}	0.119* (0.067)	-0.148 (0.120)	0.143 (0.224)
op_{t-8}	0.087 (0.066)	0.124 (0.119)	0.018 (0.222)
op_{t-9}	0.099 (0.067)	-0.304** (0.120)	0.661*** (0.223)
op_{t-10}	0.009 (0.066)	-0.066 (0.119)	-0.102 (0.222)
op_{t-11}	0.209*** (0.063)	-0.092 (0.113)	0.160 (0.211)
op_{t-12}	0.007 (0.060)	0.008 (0.107)	0.150 (0.200)
ovx_{t-1}	-0.114*** (0.032)	-0.124** (0.057)	0.035 (0.106)
ovx_{t-2}	-0.143*** (0.033)	0.162*** (0.060)	-0.122 (0.111)
ovx_{t-3}	-0.062* (0.034)	0.110* (0.062)	-0.192* (0.115)
ovx_{t-4}	-0.104*** (0.036)	0.009 (0.064)	-0.244** (0.119)
ovx_{t-5}	-0.063* (0.037)	-0.069 (0.066)	0.060 (0.123)
ovx_{t-6}	-0.050 (0.036)	-0.074 (0.064)	0.072 (0.119)
ovx_{t-7}	-0.015 (0.036)	-0.062 (0.064)	-0.089 (0.120)
ovx_{t-8}	-0.038 (0.036)	0.130** (0.065)	-0.248** (0.121)
ovx_{t-9}	-0.010 (0.034)	-0.082 (0.062)	0.273** (0.115)
ovx_{t-10}	-0.012 (0.034)	-0.079 (0.061)	0.181 (0.113)
ovx_{t-11}	-0.052 (0.034)	-0.026 (0.061)	-0.082 (0.113)
ovx_{t-12}	-0.067** (0.033)	0.108* (0.059)	-0.067 (0.111)
Constant	0.006 (0.004)	-0.004 (0.007)	0.004 (0.012)
Observations	169	169	169
RMSE	0.052	0.093	0.174

Notes:
Standard errors are in parentheses * $p < .10$, ** $p < .05$, *** $p < .01$

(b) Endogenous variables: rig_{na} op jmu

	$rig_{na,t}$	op_t	jmu
$rig_{na,t-1}$	0.109** (0.047)	-0.117** (0.047)	0.005 (0.003)
$rig_{na,t-2}$	-0.228*** (0.047)	0.022 (0.047)	7.8e-04 (0.003)
$rig_{na,t-3}$	-0.196*** (0.048)	-0.133*** (0.048)	-0.003 (0.004)
$rig_{na,t-4}$	-0.046 (0.049)	0.017 (0.049)	-4.9e-04 (0.004)
$rig_{na,t-5}$	-0.030 (0.048)	-0.034 (0.048)	-0.002 (0.004)
$rig_{na,t-6}$	-0.091* (0.047)	-0.054 (0.047)	-0.005 (0.003)
$rig_{na,t-7}$	-0.062 (0.047)	-0.013 (0.047)	-0.001 (0.003)
$rig_{na,t-8}$	0.041 (0.046)	0.091** (0.046)	-0.003 (0.003)
$rig_{na,t-9}$	-0.061 (0.045)	-0.089** (0.045)	-0.005 (0.003)
$rig_{na,t-10}$	-0.158*** (0.045)	-0.047 (0.045)	-0.001 (0.003)
$rig_{na,t-11}$	0.066 (0.045)	-0.008 (0.045)	0.003 (0.003)
$rig_{na,t-12}$	0.290*** (0.042)	-0.024 (0.042)	2.2e-04 (0.003)
op_{t-1}	0.150*** (0.050)	0.477*** (0.050)	0.008** (0.004)
op_{t-2}	0.183*** (0.057)	-0.188*** (0.057)	0.002 (0.004)
op_{t-3}	0.181*** (0.057)	0.027 (0.057)	0.003 (0.004)
op_{t-4}	0.201*** (0.057)	-0.092 (0.057)	0.002 (0.004)
op_{t-5}	0.273*** (0.058)	0.065 (0.058)	-0.002 (0.004)
op_{t-6}	0.154*** (0.059)	-0.024 (0.059)	0.006 (0.004)
op_{t-7}	0.237*** (0.059)	0.089 (0.059)	-0.003 (0.004)
op_{t-8}	0.137** (0.060)	0.037 (0.059)	0.005 (0.004)
op_{t-9}	0.155*** (0.059)	-0.012 (0.058)	-8e-04 (0.004)
op_{t-10}	0.149** (0.058)	0.100* (0.058)	0.006 (0.004)
op_{t-11}	0.153*** (0.056)	0.030 (0.056)	0.002 (0.004)
op_{t-12}	0.082 (0.053)	0.010 (0.053)	0.009** (0.004)
jmu_{t-1}	0.590 (0.684)	-4.246*** (0.683)	0.854*** (0.050)
jmu_{t-2}	-1.761* (0.921)	2.150** (0.920)	-0.055 (0.067)
jmu_{t-3}	-0.548 (0.929)	1.588* (0.927)	-0.016 (0.067)
jmu_{t-4}	-0.447 (0.931)	-2.135** (0.929)	-0.040 (0.068)
jmu_{t-5}	-0.869 (0.940)	0.384 (0.938)	0.019 (0.068)
jmu_{t-6}	0.173 (0.937)	-2.355** (0.935)	-0.048 (0.068)
jmu_{t-7}	-1.928** (0.939)	1.745* (0.937)	0.036 (0.068)
jmu_{t-8}	1.399 (0.947)	-0.892 (0.945)	-0.045 (0.069)
jmu_{t-9}	-0.871 (0.955)	1.510 (0.953)	0.046 (0.069)
jmu_{t-10}	1.237 (0.959)	-0.796 (0.957)	-0.089 (0.070)
jmu_{t-11}	0.032 (0.967)	0.074 (0.965)	0.066 (0.070)
jmu_{t-12}	-0.742 (0.757)	-0.268 (0.756)	-0.016 (0.055)
Constant	8.9e-04 (0.004)	0.002 (0.004)	5.9e-05 (2.7e-04)
Observations	401	401	401
RMSE	0.077	0.077	0.006

Notes:
Standard errors are in parentheses * $p < .10$, ** $p < .05$, *** $p < .01$

(c) Endogenous variables: rig_{na} op gepu

	$\text{rig}_{na,t}$	op_t	gepu
$\text{rig}_{na,t-1}$	0.157^{***} (0.055)	-0.133^{**} (0.058)	0.104 (0.108)
$\text{rig}_{na,t-2}$	-0.182^{***} (0.055)	0.031 (0.058)	-0.004 (0.108)
$\text{rig}_{na,t-3}$	-0.168^{***} (0.056)	-0.098^{*} (0.058)	-0.128 (0.109)
$\text{rig}_{na,t-4}$	-0.026 (0.057)	0.003 (0.060)	0.086 (0.112)
$\text{rig}_{na,t-5}$	0.023 (0.057)	0.046 (0.059)	-0.071 (0.110)
$\text{rig}_{na,t-6}$	-0.112^{**} (0.056)	-0.056 (0.058)	-0.293^{***} (0.109)
$\text{rig}_{na,t-7}$	-0.061 (0.057)	0.021 (0.059)	0.096 (0.111)
$\text{rig}_{na,t-8}$	0.052 (0.055)	0.125^{**} (0.058)	0.061 (0.108)
$\text{rig}_{na,t-9}$	-0.096^{*} (0.054)	-0.075 (0.056)	-0.172 (0.105)
$\text{rig}_{na,t-10}$	-0.184^{***} (0.054)	-0.072 (0.056)	0.290^{***} (0.104)
$\text{rig}_{na,t-11}$	0.100^{*} (0.055)	0.019 (0.057)	0.025 (0.106)
$\text{rig}_{na,t-12}$	0.297^{***} (0.050)	-0.025 (0.052)	-0.055 (0.098)
op_{t-1}	0.242^{***} (0.058)	0.486^{***} (0.060)	-0.082 (0.112)
op_{t-2}	0.239^{***} (0.065)	-0.192^{***} (0.068)	-0.099 (0.128)
op_{t-3}	0.208^{***} (0.067)	-0.012 (0.070)	0.136 (0.131)
op_{t-4}	0.269^{***} (0.067)	-0.005 (0.070)	-0.070 (0.131)
op_{t-5}	0.271^{***} (0.069)	0.058 (0.072)	0.159 (0.135)
op_{t-6}	0.185^{***} (0.070)	0.015 (0.073)	-0.003 (0.137)
op_{t-7}	0.194^{***} (0.071)	0.034 (0.073)	-0.061 (0.138)
op_{t-8}	0.083 (0.070)	-0.013 (0.073)	0.357^{***} (0.137)
op_{t-9}	0.124^{*} (0.071)	-0.070 (0.074)	-0.064 (0.138)
op_{t-10}	0.072 (0.070)	0.068 (0.072)	0.226^{*} (0.135)
op_{t-11}	0.166^{**} (0.067)	-0.018 (0.069)	0.049 (0.130)
op_{t-12}	0.065 (0.063)	-0.019 (0.066)	0.057 (0.123)
gepu_{t-1}	-0.050 (0.031)	-0.051 (0.032)	-0.230^{***} (0.060)
gepu_{t-2}	0.025 (0.032)	-0.026 (0.033)	-0.213^{***} (0.062)
gepu_{t-3}	-0.053 (0.032)	8.9e-04 (0.034)	-0.228^{***} (0.063)
gepu_{t-4}	0.013 (0.033)	-0.066^{*} (0.035)	-0.078 (0.065)
gepu_{t-5}	0.004 (0.033)	-0.028 (0.035)	-0.105 (0.065)
gepu_{t-6}	0.028 (0.033)	0.030 (0.035)	-0.025 (0.065)
gepu_{t-7}	0.001 (0.033)	-0.050 (0.035)	-0.104 (0.065)
gepu_{t-8}	0.046 (0.033)	0.033 (0.035)	-0.054 (0.065)
gepu_{t-9}	0.009 (0.033)	-0.036 (0.034)	0.049 (0.065)
gepu_{t-10}	-0.020 (0.032)	0.001 (0.034)	0.084 (0.063)
gepu_{t-11}	-0.003 (0.031)	-0.018 (0.033)	0.067 (0.061)
gepu_{t-12}	-0.020 (0.031)	-0.013 (0.032)	0.068 (0.060)
Constant	-0.004 (0.005)	0.005 (0.005)	0.006 (0.009)
Observations	292	292	292
RMSE	0.086	0.090	0.168

Notes:
Standard errors are in parentheses * $p < .10$, ** $p < .05$, *** $p < .01$

(d) Endogenous variables: rig_{na} op gtu_{op}

	$\text{rig}_{na,t}$	op_t	$\text{gtu}_{op,t}$
$\text{rig}_{na,t-1}$	0.136** (0.061)	-0.123 (0.086)	-0.034 (0.240)
$\text{rig}_{na,t-2}$	-0.187*** (0.060)	0.014 (0.084)	0.039 (0.235)
$\text{rig}_{na,t-3}$	-0.087 (0.061)	-0.158* (0.086)	-0.272 (0.239)
$\text{rig}_{na,t-4}$	-0.070 (0.061)	0.106 (0.087)	0.078 (0.241)
$\text{rig}_{na,t-5}$	-0.088 (0.059)	-0.010 (0.084)	-0.110 (0.233)
$\text{rig}_{na,t-6}$	-0.038 (0.057)	-0.050 (0.080)	-0.156 (0.224)
$\text{rig}_{na,t-7}$	-0.099* (0.056)	0.065 (0.079)	-0.002 (0.221)
$\text{rig}_{na,t-8}$	-0.009 (0.054)	0.159** (0.077)	0.207 (0.213)
$\text{rig}_{na,t-9}$	-0.156*** (0.052)	-0.017 (0.074)	-0.210 (0.206)
$\text{rig}_{na,t-10}$	-0.192*** (0.050)	-0.069 (0.071)	0.114 (0.197)
$\text{rig}_{na,t-11}$	0.062 (0.052)	0.050 (0.073)	-0.184 (0.204)
$\text{rig}_{na,t-12}$	0.382*** (0.047)	-0.075 (0.067)	0.016 (0.186)
op_{t-1}	0.191*** (0.052)	0.498*** (0.073)	-0.095 (0.203)
op_{t-2}	0.194*** (0.059)	-0.190** (0.084)	0.275 (0.234)
op_{t-3}	0.270*** (0.059)	-0.088 (0.084)	0.002 (0.234)
op_{t-4}	0.193*** (0.061)	0.080 (0.086)	0.270 (0.238)
op_{t-5}	0.263*** (0.062)	-0.017 (0.088)	0.201 (0.246)
op_{t-6}	0.141** (0.064)	-0.044 (0.090)	0.376 (0.250)
op_{t-7}	0.210*** (0.064)	0.098 (0.090)	-0.022 (0.252)
op_{t-8}	0.046 (0.065)	-0.044 (0.092)	0.391 (0.256)
op_{t-9}	0.142** (0.063)	-0.049 (0.089)	-0.106 (0.248)
op_{t-10}	0.075 (0.061)	0.040 (0.087)	0.063 (0.242)
op_{t-11}	0.189*** (0.058)	-0.059 (0.082)	0.202 (0.228)
op_{t-12}	0.093* (0.056)	-0.060 (0.079)	0.171 (0.219)
$\text{gtu}_{op,t-1}$	-0.039** (0.018)	-0.070*** (0.026)	-0.044 (0.072)
$\text{gtu}_{op,t-2}$	-0.053*** (0.019)	0.029 (0.027)	-0.132* (0.074)
$\text{gtu}_{op,t-3}$	-0.075*** (0.019)	-0.037 (0.027)	-0.009 (0.075)
$\text{gtu}_{op,t-4}$	-0.013 (0.020)	-0.036 (0.028)	-0.091 (0.079)
$\text{gtu}_{op,t-5}$	-0.043** (0.020)	-0.009 (0.029)	-0.031 (0.080)
$\text{gtu}_{op,t-6}$	-0.040** (0.020)	-0.079*** (0.028)	0.034 (0.077)
$\text{gtu}_{op,t-7}$	-0.015 (0.020)	-0.053* (0.028)	-0.092 (0.078)
$\text{gtu}_{op,t-8}$	-0.055*** (0.020)	0.005 (0.028)	0.018 (0.078)
$\text{gtu}_{op,t-9}$	0.007 (0.019)	-0.032 (0.027)	-0.066 (0.076)
$\text{gtu}_{op,t-10}$	-0.012 (0.019)	0.037 (0.027)	0.032 (0.075)
$\text{gtu}_{op,t-11}$	-0.040** (0.019)	-3.7e-04 (0.027)	-0.050 (0.075)
$\text{gtu}_{op,t-12}$	-0.050*** (0.019)	-0.016 (0.027)	0.137* (0.074)
Constant	0.005 (0.004)	0.004 (0.006)	0.010 (0.016)
Observations	209	209	209
RMSE	0.064	0.091	0.253

Notes:
Standard errors are in parentheses * $p < .10$, ** $p < .05$, *** $p < .01$

(e) Endogenous variables: rig_{na} op gtu_{oi}

	$\text{rig}_{na,t}$	op_t	$\text{gtu}_{oi,t}$
$\text{rig}_{na,t-1}$	0.126** (0.061)	-0.106 (0.087)	-0.073 (0.144)
$\text{rig}_{na,t-2}$	-0.165*** (0.060)	-0.027 (0.086)	-0.001 (0.142)
$\text{rig}_{na,t-3}$	-0.086 (0.061)	-0.135 (0.087)	-0.119 (0.144)
$\text{rig}_{na,t-4}$	-0.034 (0.061)	0.041 (0.087)	-0.019 (0.144)
$\text{rig}_{na,t-5}$	-0.073 (0.059)	-0.017 (0.085)	-0.026 (0.141)
$\text{rig}_{na,t-6}$	-0.028 (0.057)	-0.035 (0.081)	-0.153 (0.134)
$\text{rig}_{na,t-7}$	-0.069 (0.056)	0.034 (0.081)	-0.025 (0.133)
$\text{rig}_{na,t-8}$	0.006 (0.054)	0.193** (0.077)	0.168 (0.128)
$\text{rig}_{na,t-9}$	-0.166*** (0.052)	0.006 (0.075)	-0.115 (0.123)
$\text{rig}_{na,t-10}$	-0.175*** (0.050)	-0.077 (0.071)	0.016 (0.118)
$\text{rig}_{na,t-11}$	0.062 (0.051)	0.059 (0.073)	-0.119 (0.121)
$\text{rig}_{na,t-12}$	0.381*** (0.048)	-0.086 (0.068)	0.014 (0.113)
op_{t-1}	0.182*** (0.051)	0.484*** (0.073)	-0.199 (0.121)
op_{t-2}	0.172*** (0.058)	-0.185** (0.083)	0.326** (0.137)
op_{t-3}	0.297*** (0.058)	-0.066 (0.083)	-0.126 (0.138)
op_{t-4}	0.179*** (0.061)	0.045 (0.087)	0.278* (0.144)
op_{t-5}	0.269*** (0.063)	0.046 (0.090)	0.134 (0.148)
op_{t-6}	0.151** (0.063)	-0.079 (0.091)	0.226 (0.150)
op_{t-7}	0.190*** (0.064)	0.128 (0.091)	0.050 (0.151)
op_{t-8}	0.022 (0.065)	0.009 (0.093)	0.289* (0.154)
op_{t-9}	0.142** (0.063)	-0.040 (0.090)	-0.003 (0.149)
op_{t-10}	0.063 (0.061)	0.061 (0.088)	0.011 (0.146)
op_{t-11}	0.144** (0.058)	-0.039 (0.083)	0.251* (0.137)
op_{t-12}	0.096* (0.055)	-0.065 (0.079)	0.156 (0.131)
$\text{gtu}_{oi,t-1}$	-0.090*** (0.030)	-0.132*** (0.043)	-0.103 (0.072)
$\text{gtu}_{oi,t-2}$	-0.127*** (0.032)	0.073 (0.046)	-0.214*** (0.076)
$\text{gtu}_{oi,t-3}$	-0.143*** (0.034)	-0.080* (0.048)	-0.025 (0.080)
$\text{gtu}_{oi,t-4}$	-0.023 (0.037)	-0.053 (0.053)	-0.200S** (0.087)
$\text{gtu}_{oi,t-5}$	-0.077** (0.038)	-0.057 (0.054)	-0.012 (0.089)
$\text{gtu}_{oi,t-6}$	-0.044 (0.037)	-0.144*** (0.053)	-0.078 (0.087)
$\text{gtu}_{oi,t-7}$	-0.023 (0.037)	-0.145*** (0.053)	-0.120 (0.088)
$\text{gtu}_{oi,t-8}$	-0.102*** (0.038)	0.024 (0.054)	-0.063 (0.089)
$\text{gtu}_{oi,t-9}$	0.011 (0.037)	-0.045 (0.053)	-0.018 (0.088)
$\text{gtu}_{oi,t-10}$	-0.007 (0.036)	0.056 (0.052)	-0.054 (0.086)
$\text{gtu}_{oi,t-11}$	-0.105*** (0.035)	0.021 (0.051)	0.026 (0.084)
$\text{gtu}_{oi,t-12}$	-0.092*** (0.035)	-0.008 (0.050)	0.196** (0.082)
Constant	0.004 (0.004)	0.003 (0.006)	0.004 (0.009)
Observations	209	209	209
RMSE	0.062	0.089	0.148

Notes:
Standard errors are in parentheses * $p < .10$, ** $p < .05$, *** $p < .01$

(f) Endogenous variables: rig_{na} op gtu_{os}

	$\text{rig}_{na,t}$	op_t	$\text{gtu}_{os,t}$
$\text{rig}_{na,t-1}$	0.115* (0.061)	-0.101 (0.086)	0.006 (0.203)
$\text{rig}_{na,t-2}$	-0.181*** (0.059)	-0.014 (0.084)	-0.043 (0.198)
$\text{rig}_{na,t-3}$	-0.073 (0.060)	-0.140 (0.086)	-0.027 (0.202)
$\text{rig}_{na,t-4}$	-0.069 (0.060)	0.118 (0.085)	0.017 (0.201)
$\text{rig}_{na,t-5}$	-0.094 (0.059)	-0.003 (0.084)	-0.060 (0.197)
$\text{rig}_{na,t-6}$	-0.013 (0.057)	-0.054 (0.080)	-0.241 (0.189)
$\text{rig}_{na,t-7}$	-0.104* (0.056)	0.066 (0.080)	0.082 (0.188)
$\text{rig}_{na,t-8}$	-0.008 (0.054)	0.180** (0.077)	0.053 (0.182)
$\text{rig}_{na,t-9}$	-0.142*** (0.053)	-0.032 (0.075)	-0.102 (0.176)
$\text{rig}_{na,t-10}$	-0.197*** (0.050)	-0.017 (0.071)	0.115 (0.168)
$\text{rig}_{na,t-11}$	0.046 (0.052)	0.032 (0.073)	-0.233 (0.172)
$\text{rig}_{na,t-12}$	0.393*** (0.048)	-0.102 (0.067)	0.042 (0.159)
op_{t-1}	0.172*** (0.052)	0.487*** (0.074)	-0.068 (0.175)
op_{t-2}	0.205*** (0.059)	-0.178** (0.084)	0.252 (0.199)
op_{t-3}	0.274*** (0.060)	-0.077 (0.086)	-0.037 (0.202)
op_{t-4}	0.184*** (0.062)	0.058 (0.087)	0.189 (0.206)
op_{t-5}	0.256*** (0.063)	0.017 (0.089)	0.193 (0.210)
op_{t-6}	0.150** (0.064)	-0.087 (0.090)	0.178 (0.213)
op_{t-7}	0.202*** (0.064)	0.086 (0.090)	0.129 (0.213)
op_{t-8}	0.055 (0.065)	-0.008 (0.092)	0.274 (0.218)
op_{t-9}	0.153** (0.063)	-0.079 (0.090)	-0.058 (0.211)
op_{t-10}	0.067 (0.062)	0.039 (0.088)	-0.037 (0.206)
op_{t-11}	0.164*** (0.059)	-0.068 (0.083)	0.258 (0.196)
op_{t-12}	0.099* (0.056)	-0.067 (0.079)	0.146 (0.187)
$\text{gtu}_{os,t-1}$	-0.053** (0.022)	-0.094*** (0.031)	-0.018 (0.073)
$\text{gtu}_{os,t-2}$	-0.068*** (0.023)	0.042 (0.032)	-0.169** (0.076)
$\text{gtu}_{os,t-3}$	-0.095*** (0.023)	-0.044 (0.033)	-0.041 (0.078)
$\text{gtu}_{os,t-4}$	-0.021 (0.025)	-0.034 (0.035)	-0.107 (0.083)
$\text{gtu}_{os,t-5}$	-0.066*** (0.025)	-0.010 (0.036)	-0.042 (0.085)
$\text{gtu}_{os,t-6}$	-0.051** (0.025)	-0.082** (0.035)	0.010 (0.083)
$\text{gtu}_{os,t-7}$	-0.027 (0.025)	-0.088** (0.036)	-0.099 (0.084)
$\text{gtu}_{os,t-8}$	-0.064** (0.025)	0.025 (0.036)	0.067 (0.084)
$\text{gtu}_{os,t-9}$	0.023 (0.025)	-0.034 (0.035)	-0.102 (0.083)
$\text{gtu}_{os,t-10}$	-0.015 (0.025)	0.032 (0.035)	-0.029 (0.083)
$\text{gtu}_{os,t-11}$	-0.060** (0.024)	0.007 (0.034)	-0.005 (0.081)
$\text{gtu}_{os,t-12}$	-0.061*** (0.024)	-0.001 (0.033)	0.127 (0.079)
Constant	0.004 (0.004)	0.003 (0.006)	0.006 (0.014)
Observations	209	209	209
RMSE	0.064	0.090	0.213

Notes:
Standard errors are in parentheses * $p < .10$, ** $p < .05$, *** $p < .01$

(g) Endogenous variables: rig_{na} op gtu_{od}

	$\text{rig}_{na,t}$	op_t	$\text{gtu}_{od,t}$
$\text{rig}_{na,t-1}$	0.118* (0.062)	-0.081 (0.092)	0.057 (0.199)
$\text{rig}_{na,t-2}$	-0.169*** (0.060)	0.036 (0.089)	0.145 (0.194)
$\text{rig}_{na,t-3}$	-0.078 (0.061)	-0.089 (0.090)	0.043 (0.196)
$\text{rig}_{na,t-4}$	-0.067 (0.060)	0.127 (0.089)	-0.064 (0.194)
$\text{rig}_{na,t-5}$	-0.102* (0.058)	-0.039 (0.086)	-0.032 (0.188)
$\text{rig}_{na,t-6}$	-0.057 (0.056)	-0.049 (0.083)	-0.087 (0.181)
$\text{rig}_{na,t-7}$	-0.105* (0.056)	0.064 (0.083)	-0.069 (0.180)
$\text{rig}_{na,t-8}$	0.014 (0.054)	0.159** (0.080)	0.246 (0.174)
$\text{rig}_{na,t-9}$	-0.144*** (0.052)	0.037 (0.077)	-0.186 (0.167)
$\text{rig}_{na,t-10}$	-0.184*** (0.049)	-0.071 (0.073)	0.049 (0.159)
$\text{rig}_{na,t-11}$	0.040 (0.050)	0.035 (0.075)	-0.181 (0.163)
$\text{rig}_{na,t-12}$	0.364*** (0.046)	-0.085 (0.069)	0.086 (0.150)
op_{t-1}	0.147*** (0.051)	0.467*** (0.075)	-0.114 (0.163)
op_{t-2}	0.189*** (0.057)	-0.148* (0.084)	0.211 (0.183)
op_{t-3}	0.274*** (0.057)	-0.058 (0.084)	-0.048 (0.183)
op_{t-4}	0.180*** (0.059)	0.031 (0.087)	0.116 (0.190)
op_{t-5}	0.234*** (0.060)	-0.034 (0.089)	0.153 (0.194)
op_{t-6}	0.128** (0.060)	-0.091 (0.090)	0.226 (0.195)
op_{t-7}	0.203*** (0.061)	0.022 (0.091)	0.135 (0.197)
op_{t-8}	0.065 (0.062)	-0.049 (0.092)	0.242 (0.200)
op_{t-9}	0.182*** (0.060)	-0.084 (0.090)	-0.095 (0.195)
op_{t-10}	0.071 (0.060)	0.012 (0.089)	-0.002 (0.193)
op_{t-11}	0.168*** (0.057)	-0.091 (0.085)	0.240 (0.185)
op_{t-12}	0.105* (0.054)	-0.051 (0.080)	0.080 (0.173)
$\text{gtu}_{od,t-1}$	-0.072*** (0.023)	-0.105*** (0.034)	-0.144* (0.074)
$\text{gtu}_{od,t-2}$	-0.101*** (0.025)	0.040 (0.037)	-0.121 (0.080)
$\text{gtu}_{od,t-3}$	-0.112*** (0.025)	5e-04 (0.038)	-0.016 (0.082)
$\text{gtu}_{od,t-4}$	-0.007 (0.027)	-0.002 (0.041)	-0.067 (0.089)
$\text{gtu}_{od,t-5}$	-0.079*** (0.028)	0.021 (0.041)	0.037 (0.090)
$\text{gtu}_{od,t-6}$	-0.068** (0.028)	-0.071* (0.041)	0.099 (0.089)
$\text{gtu}_{od,t-7}$	-0.047* (0.028)	-0.084** (0.041)	0.026 (0.089)
$\text{gtu}_{od,t-8}$	-0.080*** (0.028)	-0.025 (0.041)	0.125 (0.090)
$\text{gtu}_{od,t-9}$	0.018 (0.028)	-0.019 (0.042)	0.025 (0.090)
$\text{gtu}_{od,t-10}$	0.009 (0.028)	0.035 (0.041)	-0.055 (0.089)
$\text{gtu}_{od,t-11}$	-0.048* (0.027)	-0.025 (0.041)	0.043 (0.088)
$\text{gtu}_{od,t-12}$	-0.073*** (0.026)	6e-04 (0.039)	0.131 (0.084)
Constant	0.005 (0.004)	0.003 (0.006)	0.006 (0.013)
Observations	209	209	209
RMSE	0.062	0.092	0.199

Notes:
Standard errors are in parentheses * $p < .10$, ** $p < .05$, *** $p < .01$

(h) Endogenous variables: rig_{na} op gtu_{oms}

	$\text{rig}_{na,t}$	op_{t}	$\text{gtu}_{oms,t}$
$\text{rig}_{na,t-1}$	0.119* (0.061)	-0.102 (0.089)	-0.025 (0.174)
$\text{rig}_{na,t-2}$	-0.178*** (0.060)	-0.001 (0.087)	0.050 (0.171)
$\text{rig}_{na,t-3}$	-0.085 (0.060)	-0.122 (0.088)	-0.036 (0.173)
$\text{rig}_{na,t-4}$	-0.059 (0.060)	0.086 (0.088)	-0.025 (0.172)
$\text{rig}_{na,t-5}$	-0.095 (0.058)	-0.020 (0.085)	0.012 (0.167)
$\text{rig}_{na,t-6}$	-0.033 (0.056)	-0.047 (0.082)	-0.164 (0.160)
$\text{rig}_{na,t-7}$	-0.093* (0.055)	0.054 (0.081)	0.003 (0.159)
$\text{rig}_{na,t-8}$	0.002 (0.054)	0.178** (0.078)	0.161 (0.154)
$\text{rig}_{na,t-9}$	-0.154*** (0.051)	0.009 (0.075)	-0.142 (0.147)
$\text{rig}_{na,t-10}$	-0.180*** (0.049)	-0.063 (0.072)	0.057 (0.140)
$\text{rig}_{na,t-11}$	0.048 (0.050)	0.046 (0.074)	-0.167 (0.144)
$\text{rig}_{na,t-12}$	0.381*** (0.047)	-0.093 (0.068)	0.040 (0.134)
op_{t-1}	0.158*** (0.051)	0.475*** (0.075)	-0.115 (0.147)
op_{t-2}	0.177*** (0.058)	-0.162* (0.084)	0.265 (0.165)
op_{t-3}	0.285*** (0.058)	-0.072 (0.085)	-0.074 (0.166)
op_{t-4}	0.183*** (0.060)	0.041 (0.088)	0.215 (0.172)
op_{t-5}	0.247*** (0.062)	0.011 (0.090)	0.170 (0.176)
op_{t-6}	0.153** (0.062)	-0.092 (0.090)	0.184 (0.177)
op_{t-7}	0.204*** (0.062)	0.076 (0.091)	0.010 (0.179)
op_{t-8}	0.048 (0.064)	-0.009 (0.093)	0.252 (0.182)
op_{t-9}	0.171*** (0.062)	-0.063 (0.090)	-0.063 (0.177)
op_{t-10}	0.071 (0.061)	0.045 (0.089)	-0.020 (0.174)
op_{t-11}	0.152*** (0.058)	-0.060 (0.084)	0.257 (0.165)
op_{t-12}	0.103* (0.054)	-0.053 (0.079)	0.115 (0.156)
$\text{gtu}_{oms,t-1}$	-0.078*** (0.026)	-0.118*** (0.037)	-0.081 (0.073)
$\text{gtu}_{oms,t-2}$	-0.108*** (0.027)	0.056 (0.040)	-0.166** (0.078)
$\text{gtu}_{oms,t-3}$	-0.128*** (0.028)	-0.041 (0.041)	-0.035 (0.081)
$\text{gtu}_{oms,t-4}$	-0.015 (0.031)	-0.037 (0.045)	-0.115 (0.088)
$\text{gtu}_{oms,t-5}$	-0.087*** (0.031)	-0.013 (0.046)	0.014 (0.090)
$\text{gtu}_{oms,t-6}$	-0.059* (0.031)	-0.107** (0.045)	0.021 (0.088)
$\text{gtu}_{oms,t-7}$	-0.032 (0.031)	-0.116** (0.045)	-0.060 (0.089)
$\text{gtu}_{oms,t-8}$	-0.091*** (0.031)	0.006 (0.046)	0.075 (0.089)
$\text{gtu}_{oms,t-9}$	0.017 (0.031)	-0.030 (0.045)	-0.026 (0.089)
$\text{gtu}_{oms,t-10}$	-0.004 (0.030)	0.046 (0.045)	-0.040 (0.087)
$\text{gtu}_{oms,t-11}$	-0.074** (0.030)	0.003 (0.044)	0.035 (0.086)
$\text{gtu}_{oms,t-12}$	-0.085*** (0.029)	0.006 (0.043)	0.164** (0.083)
Constant	0.005 (0.004)	0.003 (0.006)	0.005 (0.011)
Observations	209	209	209
RMSE	0.062	0.090	0.177

Notes:
Standard errors are in parentheses * $p < .10$, ** $p < .05$, *** $p < .01$

B.4 Autocorrelation test

In order to detect whether or not the model is misspecified, the autocorrelation test is conducted. The test is conducted by regressing the residual of the VAR model (from Equation 4.7) into the independent variables and the lagged of the residuals that is so-called auxiliary regression.

$$\hat{e_t} = \gamma_0 + \rho_1 \hat{e_{t-1}} + \rho_2 \hat{e_{t-2}} + ... + \rho_q \hat{e_{t-q}} + \gamma_1 x_t + u_t \tag{B.2}$$

The presence of autocorrelation in the error term causes the OLS standard error and t-statistic are no longer valid and the estimated OLS coefficients and variances will be biased and inconsistent due to the presence of lagged dependent variables. Hence, it requires a corrected t-test by applying a Lagrange Multiplier (LM) t-statistic as to the following (Søren, 2003).

$$LM = \left(T - pk - m - p - \frac{1}{2}\right) log \left(\frac{\hat{\Omega}}{\tilde{\Omega}}\right) \tag{B.3}$$

T is the total observations, p^2 is the degree of freedom, k is the lag length, m is the number of restrictions, $\hat{\Omega}$ denotes variance estimate from VAR equation, and $\tilde{\Omega}$ denotes variance estimate from auxiliary regression. The t-statistic is distributed as χ^2 distribution with the degree of freedom p^2. The null hypothesis is of no serial correlation at lag order q; $H_0 : \rho_1 = 0, \rho_2 = 0, ..., \rho_q = 0$ with the critical values as described in Edgerton and Shukur (1999). Table B.5 provides the information about the autocorrelation LM test of each VAR model using lag length determined in Appendix B.1.

Table B.5 shows that the null hypothesis is rejected at 5% significance level at the specified lag based on Table B.1 so that there is no autocorrelation in all VAR specification.

Table B.5: Autocorrelation test

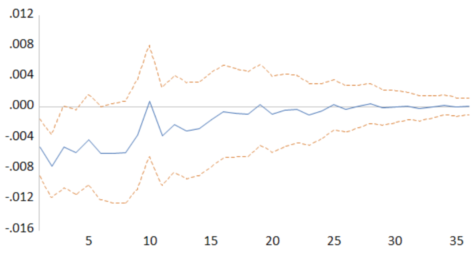
VAR endogenous variables	LR stat	Prob
rig op ovx	6.991	0.638
rig op jmu	10.175	0.337
rig op gepu	7.963	0.538
rig op gtu_{op}	9.761	0.370
rig op gtu_{oi}	12.210	0.202
rig op gtu_{os}	5.846	0.755
rig op gtu_{od}	14.418	0.108
rig op gtu_{oms}	5.007	0.834
rig_{na} op ovx	9.829	0.365
rig_{na} op jmu	14.627	0.102
rig_{na} op gepu	7.869	0.547
rig_{na} op gtu_{op}	5.389	0.799
rig_{na} op gtu_{oi}	3.884	0.919
rig_{na} op gtu_{os}	7.604	0.575
rig_{na} op gtu_{od}	15.922	0.069
rig_{na} op gtu_{oms}	7.537	0.581

B.5 Robustness test

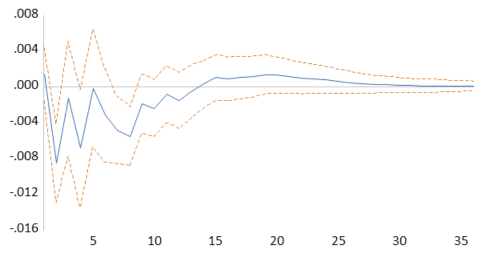
Robustness test is carried out to ensure the consistency of the empirical results.

B.5.1 Reordering variables

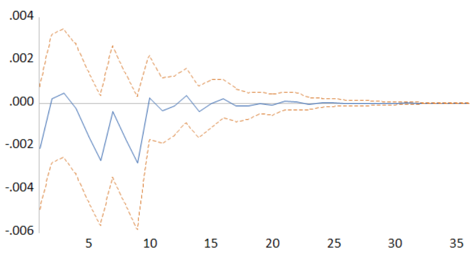
The first robustness test carries out reordering the variables to check whether contemporaneous response between variables in the impulse response functions are consistent. In this robustness test, this study does the reordering variables into oil rig count, uncertainty index, and crude oil price. The rig count is assumed not to have a contemporaneous response to the uncertainty and crude oil price, and uncertainty responds to rig count but does not have a contemporaneous response to the oil price. Meanwhile, the crude oil price responds contemporaneously to rig count and uncertainty shocks. Figure B.3 - B.6 show that all impulse responses are consistent with the order in the base empirical model. Exploration activities and crude oil price have a significant negative response to the GTU shocks.



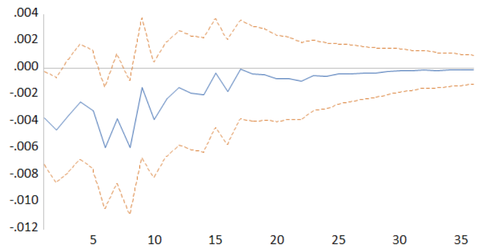
(a) Rig count response to OVX shock



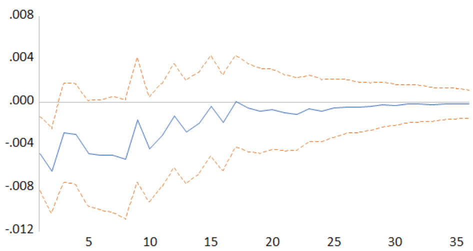
(b) Rig count response to JMU shock



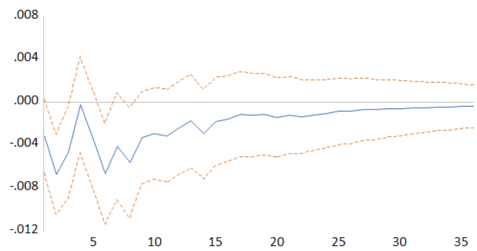
(c) Rig count response to GEPU shock



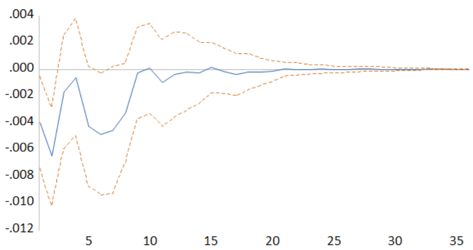
(d) Rig count response to GTU oil price shock



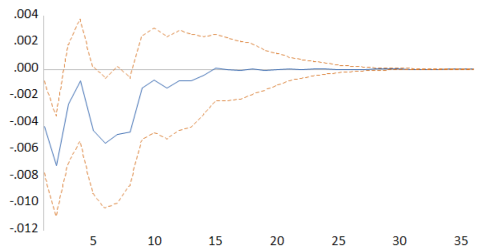
(e) Rig count response to GTU oil investment shock



(f) Rig count response to GTU oil supply shock

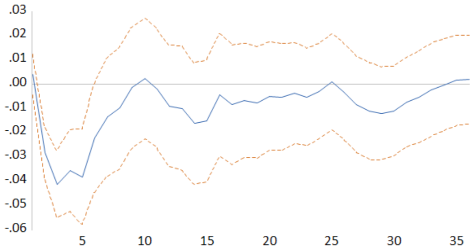


(g) Rig count response to GTU oil demand shock

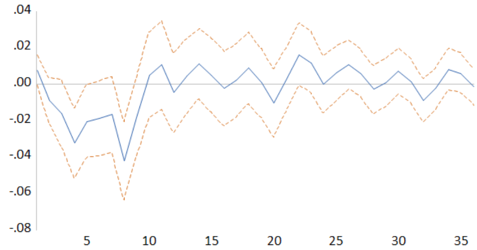


(h) Rig count response to GTU oil market specific shock

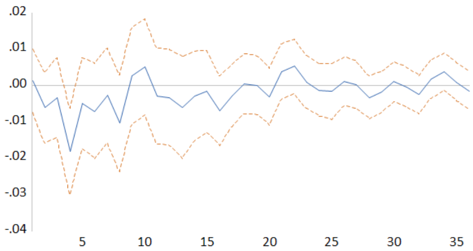
Figure B.3: World oil exploration activity responses to various uncertainty shocks (reordering: uncertainty - rig count - oil price)



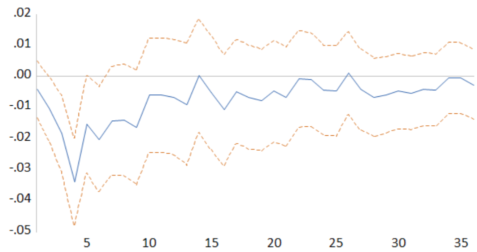
(a) Rig count response to OVX shock



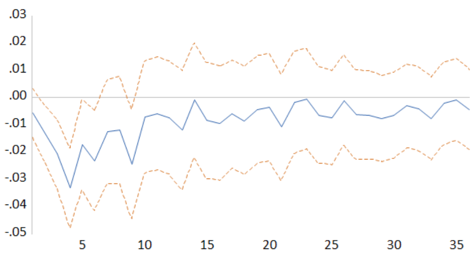
(b) Rig count response to JMU shock



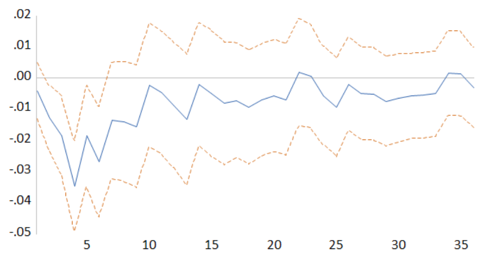
(c) Rig count response to GEPU shock



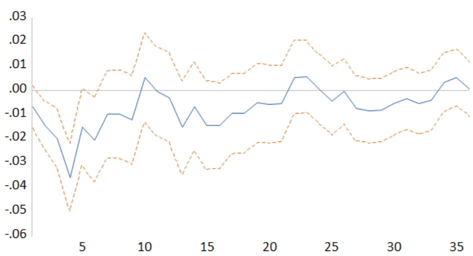
(d) Rig count response to GTU oil price shock



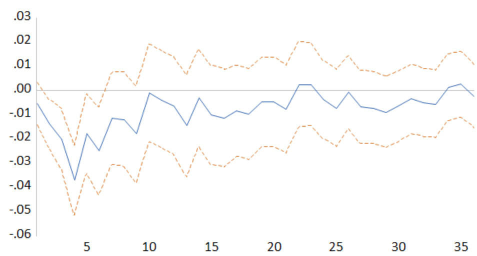
(e) Rig count response to GTU oil investment shock



(f) Rig count response to GTU oil supply shock

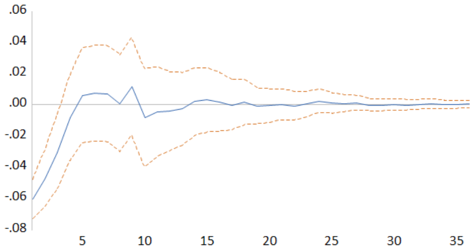


(g) Rig count response to GTU oil demand shock

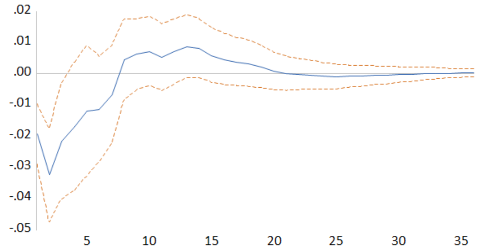


(h) Rig count response to GTU oil market specific shock

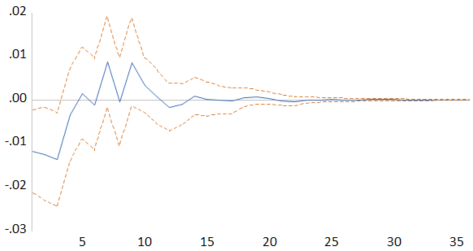
Figure B.4: North America oil exploration activity responses to various uncertainty shocks (reordering: uncertainty - rig count - oil price)



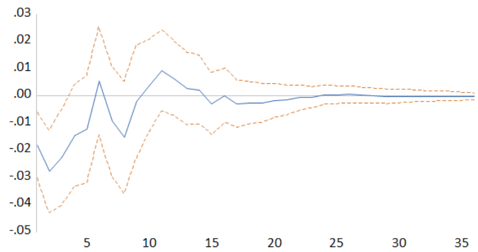
(a) Oil price response to OVX shock



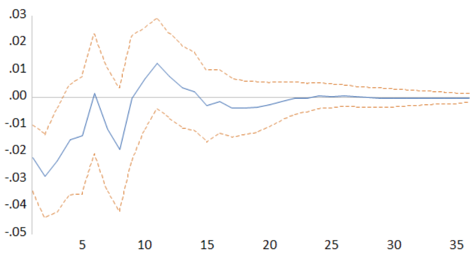
(b) Oil price response to JMU shock



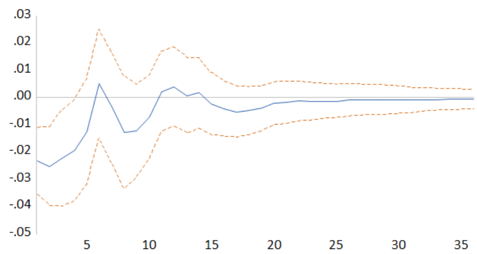
(c) Oil price response to GEPU shock



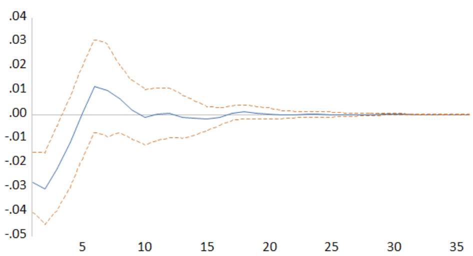
(d) Oil price response to GTU oil price shock



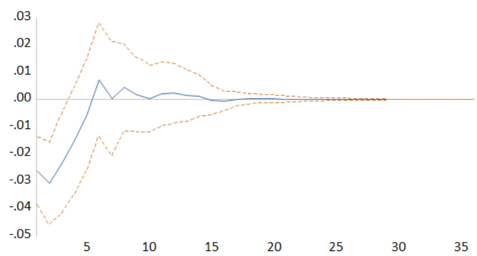
(e) Oil price response to GTU oil investment shock



(f) Oil price response to GTU oil supply shock

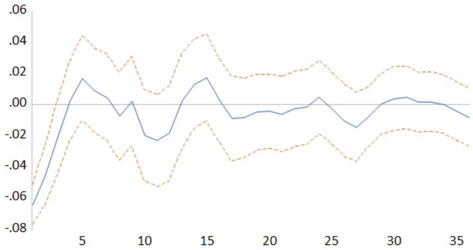


(g) Oil price response to GTU oil demand shock

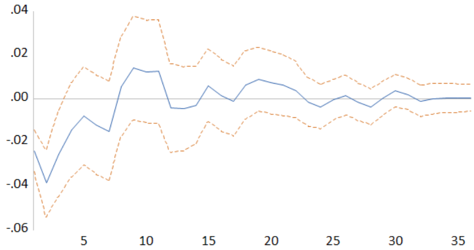


(h) Oil price response to GTU oil market specific shock

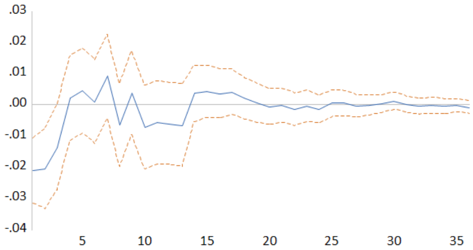
Figure B.5: Crude oil price responses to various uncertainty shocks (reordering: uncertainty - rig count - oil price) - world exploration activity model



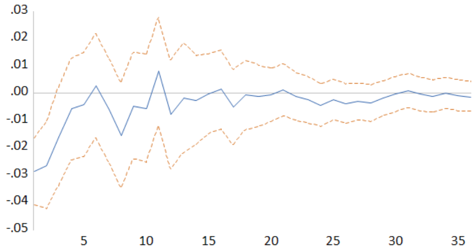
(a) Oil price response to OVX shock



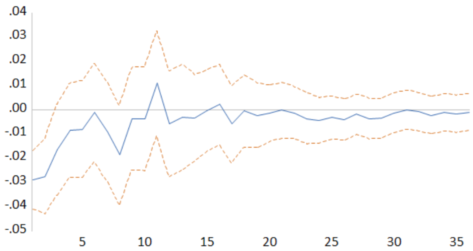
(b) Oil price response to JMU shock



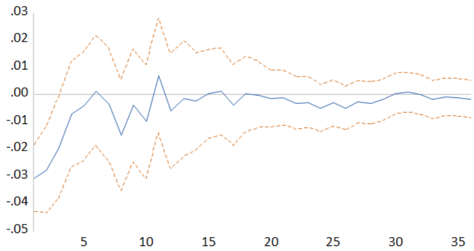
(c) Oil price response to GEPU shock



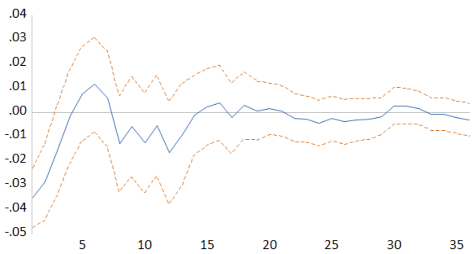
(d) Oil price response to GTU oil price shock



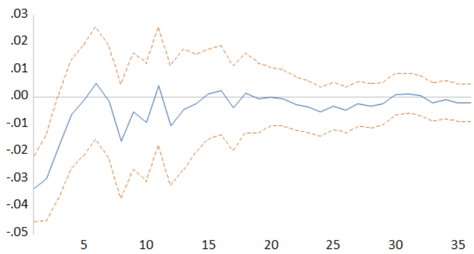
(e) Oil price response to GTU oil investment shock



(f) Oil price response to GTU oil supply shock



(g) Oil price response to GTU oil demand shock

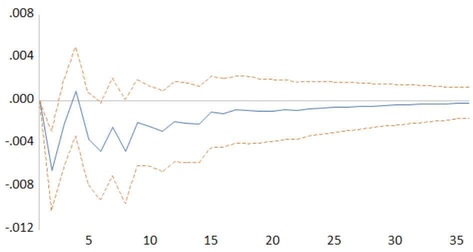


(h) Oil price response to GTU oil market specific shock

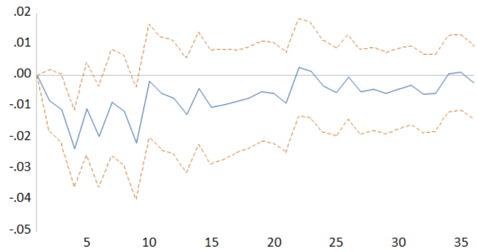
Figure B.6: Crude oil price responses to various uncertainty shocks (reordering: uncertainty - rig count - oil price) - North America exploration activity model

B.5.2 Brent oil price

The robustness test also applies brent price as the proxy of crude oil price. Unanticipated shock in GTU also gives significant negative responses on exploration activities and crude oil price.

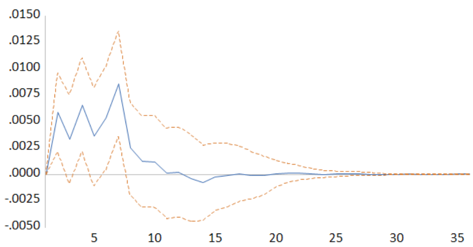


(a) World exploration activity response to GTU oil market specific shock

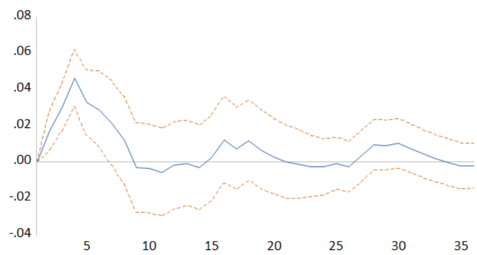


(b) North America exploration activity response to GTU oil market specific shock

Figure B.7: Rig count response to GTU oil market specific shock (Brent price model)

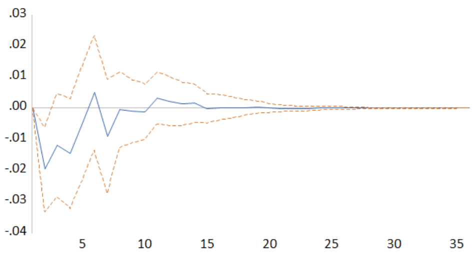


(a) World exploration activity model

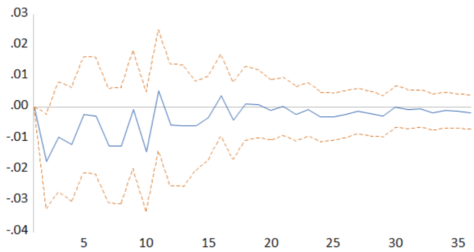


(b) North America exploration activity model

Figure B.8: Rig count response to Brent price shock



(a) World exploration activity model



(b) North America exploration activity model

Figure B.9: Brent price response to GTU oil market specific shock

Bibliography

Abraham, B. and Ledolter, J. (1983), *Statistical methods for forecasting*, Vol. 234, Wiley, New York.

Afkhami, M., Cormack, L. and Ghoddusi, H. (2017), 'Google search keywords that best predict energy price volatility', *Energy Economics* **67**, 17 – 27.
URL: *http://www.sciencedirect.com/science/article/pii/S0140988317302517*

Ahir, H., Bloom, N. and Furceri, D. (2018), 'The world uncertainty index'.
URL: *https://ssrn.com/abstract=3275033*

Ahmadi, M., Manera, M. and Sadeghzadeh, M. (2019), 'The investment-uncertainty relationship in the oil and gas industry', *Resources Policy* **63**, 101439.
URL: *http://www.sciencedirect.com/science/article/pii/S0301420719303769*

Aloui, R., Gupta, R. and Miller, S. M. (2016), 'Uncertainty and crude oil returns', *Energy Economics* **55**, 92 – 100.
URL: *http://www.sciencedirect.com/science/article/pii/S0140988316000414*

Alquist, R. and Kilian, L. (2010), 'What do we learn from the price of crude oil futures?', *Journal of Applied Econometrics* **25**(4), 539–573.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.1159*

Anderson, S. T., Kellogg, R. and Salant, S. W. (2014), Hotelling under pressure, Technical report, National Bureau of Economic Research. Copyright - National Bureau of Economic Research, Inc. Jul 2014; Last updated -

2015-12-18.
URL: *https://search-proquest-com.dbgw.lis.curtin.edu.au/docview/1687918366?accountid=10382*

Anderson, S. T., Kellogg, R. and Salant, S. W. (2018), 'Hotelling under pressure', *Journal of Political Economy* **126**(3), 984–1026.
URL: *https://doi.org/10.1086/697203*

Andrews, D. W. K. (1993), 'Tests for parameter instability and structural change with unknown change point', *Econometrica* **61**(4), 821–856.
URL: *http://www.jstor.org/stable/2951764*

Antonakakis, N., Babalos, V. and Kyei, C. (2016), 'Predictability of sustainable investments and the role of uncertainty: evidence from a non-parametric causality-in-quantiles test', *Applied Economics* **48**(48), 4655–4665.
URL: *https://doi.org/10.1080/00036846.2016.1161724*

Antonakakis, N., Chatziantoniou, I. and Filis, G. (2014), 'Dynamic spillovers of oil price shocks and economic policy uncertainty', *Energy Economics* **44**, 433 – 447.
URL: *http://www.sciencedirect.com/science/article/pii/S0140988314001157*

Apergis, N., Aslan, A., Aye, G. C. and Gupta, R. (2015), 'The asymmetric effect of oil price on growth across us states', *Energy Exploration & Exploitation* **33**(4), 575–590.
URL: *https://doi.org/10.1260/0144-5987.33.4.575*

Apergis, N. and Miller, S. M. (2009), 'Do structural oil-market shocks affect stock prices?', *Energy Economics* **31**(4), 569 – 575.
URL: *http://www.sciencedirect.com/science/article/pii/S0140988309000371*

Arps, J. J. and Roberts, T. G. (1958), 'Economics of drilling for cretaceous oil on east flank of denver-julesburg basin', *AAPG Bulletin* **42**(11), 2549–2566.

Arps, J., Mortada, M., Smith, A. et al. (1971), 'Relationship between proved reserves and exploratory effort', *Journal of Petroleum Technology* **23**(06), 671–675.

Bachmann, R., Elstner, S. and Sims, E. R. (2013), 'Uncertainty and economic activity: Evidence from business survey data', *American Economic Journal: Macroeconomics* **5**(2), 217–49.
URL: *https://www.aeaweb.org/articles?id=10.1257/mac.5.2.217*

Baker, M. and Wurgler, J. (2007), 'Investor sentiment in the stock market', *Journal of Economic Perspectives* **21**(2), 129–152.
URL: *https://www.aeaweb.org/articles?id=10.1257/jep.21.2.129*

Baker, S., Bloom, N. and Davis, S. (2016), 'Measuring economic policy uncertainty', *The Quarterly Journal of Economics* **131**(1-–12), 1593–1636.

Banerjee, P., Arčabić, V. and Lee, H. (2017), 'Fourier adl cointegration test to approximate smooth breaks with new evidence from crude oil market', *Economic Modelling* **67**, 114 – 124.
URL: *http://www.sciencedirect.com/science/article/pii/S0264999316306393*

Barouch, E. and Kaufman, G. M. (1976), 'Oil and gas discovery modelled as sampling proportional to random size'.

Basu, S. and Bundick, B. (2017), 'Uncertainty shocks in a model of effective demand', *Econometrica* **85**(3), 937–958.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA13960*

Baumeister, C. and Hamilton, J. D. (2019), 'Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks', *American Economic Review* **109**(5), 1873–1910.

Baumeister, C. and Kilian, L. (2016*a*), 'Forty years of oil price fluctuations: Why the price of oil may still surprise us', *Journal of Economic Perspectives* **30**(1), 139–60.

Baumeister, C. and Kilian, L. (2016*b*), 'Understanding the decline in the price of oil since june 2014', *Journal of the Association of Environmental and Resource Economists* **3**(1), 131–158. doi: 10.1086/684160; 01.

Baumeister, C. and Peersman, G. (2013), 'The role of time-varying price elasticities in accounting for volatility changes in the crude oil market', *Journal of Applied Econometrics* **28**(7), 1087–1109.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.2283*

Beall, A. O. (1976), 'Dynamics of petroleum industry investment in the north sea'.

Becker, R., Enders, W. and Hurn, S. (2004), 'A general test for time dependence in parameters', *Journal of Applied Econometrics* **19**(7), 899–906.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.751*

Bekiros, S., Gupta, R. and Paccagnini, A. (2015), 'Oil price forecastability and economic uncertainty', *Economics Letters* **132**, 125 – 128.
URL: *http://www.sciencedirect.com/science/article/pii/S0165176515001743*

Bilgin, M. H., Demir, E., Gozgor, G., Karabulut, G. and Kaya, H. (2019), 'A novel index of macroeconomic uncertainty for turkey based on google-trends', *Economics Letters* **184**, 108601.
URL: *http://www.sciencedirect.com/science/article/pii/S0165176519302964*

Bloom, N. (2009), 'The impact of uncertainty shocks', *econometrica* **77**(3), 623–685.

Bohling, G. C. and Davis, J. C. (1993), 'A fortran program for monte carlo simulation of oil-field discovery sequences', *Computers & Geosciences* **19**(10), 1529–1543.
URL: *https://www.sciencedirect.com/science/article/pii/009830049390066E*

Bontempi, M. E., Golinelli, R. and Squadrani, M. (2016), 'A new index of uncertainty based on internet searches: A friend or foe of other indicators?',

Quaderni - Working Paper DSE N° 1062 .
URL: *https://ssrn.com/abstract=2746346*

Broock, W. A., Scheinkman, J. A., Dechert, W. D. and LeBaron, B. (1996), 'A test for independence based on the correlation dimension', *Econometric Reviews* **15**(3), 197–235.
URL: *https://doi.org/10.1080/07474939608800353*

Caldara, D. and Iacoviello, M. (2018), 'Measuring geopolitical risk'.
URL: *https://ssrn.com/abstract=3117773*

Campos, I., Cortazar, G. and Reyes, T. (2017), 'Modeling and predicting oil vix: Internet search volume versus traditional mariables', *Energy Economics* **66**, 194 – 204.
URL: *http://www.sciencedirect.com/science/article/pii/S0140988317302037*

Casella, G. and Berger, R. L. (2021), *Statistical inference*, Cengage Learning.

Castelnuovo, E. and Tran, T. (2017), 'Google it up! a google trends-based uncertainty index for the united states and australia', *Economics Letters* **161**, 149–153.
URL: *http://www.sciencedirect.com/science/article/pii/S0165176517303993*

CBOE (2020), 'Cboe crude oil etf volatility index (ovx)'.
URL: *http://www.cboe.com/products/vix-index-volatility/volatility-on-etfs/cboe-crude-oil-etf-volatility-index-ovx*

Chen, F. and Linn, S. C. (2017), 'Investment and operating choice: Oil and natural gas futures prices and drilling activity', *Energy Economics* **66**, 54 – 68.
URL: *http://www.sciencedirect.com/science/article/pii/S014098831730155X*

Chen, Y.-F. and Mu, X. (2021), 'Asymmetric volatility in commodity markets', *Journal of Commodity Markets* **22**, 100139.
URL: *https://www.sciencedirect.com/science/article/pii/S2405851320300167*

Chu, C.-S. J., Hornik, K. and Kuan, C.-M. (1995), 'The moving-estimates test for parameter stability', *Econometric Theory* **11**(4), 699–720.

Cleveland, C. J. and Kaufmann, R. K. (1991), 'Forecasting ultimate oil recovery and its rate of production: Incorporating economic forces into the models of m. king hubbert', *The Energy Journal* **12**(2).

Cologni, A. and Manera, M. (2014), 'On the economic determinants of oil production: Theoretical analysis and empirical evidence for small exporting countries', *Energy Economics* **44**, 68 – 79.
URL: *http://www.sciencedirect.com/science/article/pii/S0140988314000668*

Cortazar, G., Schwartz, E. S. and Casassus, J. (2003), Chapter 9 - optimal exploration investments under price and geological—technical uncertainty: a real options model, *in* D. A. Paxson, ed., 'Real R & D Options', Quantitative Finance, Butterworth-Heinemann, Oxford, pp. 149 – 165.
URL: *http://www.sciencedirect.com/science/article/pii/B9780750653329500094*

Cullen, A. C. and Frey, C. H. (1999), *Probabilistic techniques in exposure assessment: a handbook for dealing with variability and uncertainty in models and inputs*, Springer Science & Business Media.

Da, Z., Engelberg, J. and Gai, P. (2011), 'In search of attention', *The Journal of Finance* **66**(5), 1461–1499.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.2011.01679.x*

Dahl, C. and Duggan, T. E. (1998), 'Survey of price elasticities from economic exploration models of us oil and gas supply', *Journal of Energy Finance & Development* **3**(2), 129–169.
URL: *https://www.sciencedirect.com/science/article/pii/S1085744399800726*

Davis, S. J. (2016), An index of global economic policy uncertainty, Working Paper 22740, National Bureau of Economic Research.
URL: *http://www.nber.org/papers/w22740*

Dees, S., Gasteuil, A., Kaufmann, R. and Mann, M. (2008), Assessing the factors behind oil price changes, Working Paper Series 855, European Central Bank.
URL: *https://EconPapers.repec.org/RePEc:ecb:ecbwps:2008855*

Degiannakis, S., Filis, G. and Panagiotakopoulou, S. (2018), 'Oil price shocks and uncertainty: How stable is their relationship over time?', *Economic Modelling* **72**, 42 – 53.
URL: *http://www.sciencedirect.com/science/article/pii/S0264999317310374*

Delignette-Muller, M. L., Dutang, C. et al. (2015), 'fitdistrplus: An r package for fitting distributions', *Journal of statistical software* **64**(4), 1–34.

Dickey, D. A. and Fuller, W. A. (1979), 'Distribution of the estimators for autoregressive time series with a unit root', *Journal of the American statistical association* **74**(366a), 427–431.

Dixit, A. K. and Pindyck, R. S. (1994), *Investment under uncertainty*, Princeton University Press.

Dunne, T. and Mu, X. (2010), 'Investment spikes and uncertainty in the petroleum refining industry', *The Journal of Industrial Economics* **58**(1), 190–213.

Dzielinski, M. (2012), 'Measuring economic uncertainty and its impact on the stock market', *Finance Research Letters* **9**(3), 167 – 175.
URL: *http://www.sciencedirect.com/science/article/pii/S1544612311000468*

Dées, S., Karadeloglou, P., Kaufmann, R. K. and Sánchez, M. (2007), 'Modelling the world oil market: Assessment of a quarterly econometric model', *Energy*

Policy **35**(1), 178 – 191.
URL: *http://www.sciencedirect.com/science/article/pii/S0301421505002983*

D'Agostino, R. B. and Stephens, M. A. (1986), 'Tests based on edf statistics'.

Eckbo, P. L., Jacoby, H. D. and Smith, J. L. (1978), 'Oil supply forecasting: A disaggregated process approach', *The Bell Journal of Economics* **9**(1), 218–235.
URL: *http://www.jstor.org/stable/3003622*

Edgerton, D. and Shukur, G. (1999), 'Testing autocorrelation in a system perspective testing autocorrelation', *Econometric Reviews* **18**(4), 343–386.
URL: *https://doi.org/10.1080/07474939908800351*

EIA (2019), 'Tight oil development will continue to drive future u.s. crude oil production'.
URL: *https://www.eia.gov/todayinenergy/detail.php?id=38852*

EIA (2021), 'International petroleum and other liquids'.
URL: *https://www.eia.gov/international/data/world/total-energy*

Elder, J. and Serletis, A. (2010), 'Oil price uncertainty', *Journal of Money, Credit and Banking* **42**(6), 1137–1159.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1538-4616.2010.00323.x*

Enders, W. (2015), *Applied econometric time series*, fourth edn, John Wiley & Sons, Inc.

Enders, W. and Lee, J. (2012), 'A unit root test using a fourier series to approximate smooth breaks', *Oxford bulletin of Economics and Statistics* **74**(4), 574 – 599.
URL: *https://EconPapers.repec.org/RePEc:bla:obuest:v:74:y:2012:i:4:p:574-599*

Engle, R. F. and Granger, C. W. J. (1987), 'Co-integration and error correction: Representation, estimation, and testing', *Econometrica (1986-1998)* **55**(2), 251. Copyright - Copyright Econometric Society Mar 1987; Last updated - 2011-10-12; CODEN - ECMTA7.
URL: *https://search-proquest-com.dbgw.lis.curtin.edu.au/docview/214870727?accountid=10382*

EViews (2019), 'Eviews 11 help user's guide - multiple equation analysis - vector autoregression and error correction models - views and procs of a var'.
URL: *http://www.eviews.com/help/helpintro.html#page/content%2FVAR-Views_and_Procs_of_a_VAR.html%23ww78271*

Ewing, B. (2017), 'Discoveries of proved reserves and the influence of oil price and interest rate', *Energy Sources, Part B: Economics, Planning and Policy* **12**(5), 452–459.

Fan, Y. and Xu, J.-H. (2011), 'What has driven oil prices since 2000? a structural change perspective', *Energy Economics* **33**(6), 1082 – 1094.
URL: *http://www.sciencedirect.com/science/article/pii/S0140988311001228*

Farzin, Y. (2001), 'The impact of oil price on additions to us proven reserves', *Resource and Energy Economics* **23**(3), 271–292.
URL: *https://www.sciencedirect.com/science/article/pii/S0928765501000409*

Fattouh, B., Kilian, L. and Mahadeva, L. (2013), 'The role of speculation in oil markets: What have we learned so far?', *The Energy Journal* **34**(3).

Fiévet, L., Forró, Z., Cauwels, P. and Sornette, D. (2015), 'A general improved methodology to forecasting future oil production: Application to the uk and norway'. ID: 271090.

Forbes, K. F. and Zampelli, E. M. (2002), 'Technology and the exploratory success rate in the u.s. onshore', *The Quarterly Review of Economics and*

Finance **42**(2), 319–334.
URL: *https://www.sciencedirect.com/science/article/pii/S106297690200128X*

Gronwald, M. (2008), 'Large oil shocks and the us economy: Infrequent incidents with large effects', *The Energy Journal* **29**(1).
URL: *http://www.iaee.org/en/publications/ejarticle.aspx?id=2245*

Gronwald, M. (2016), 'Explosive oil prices', *Energy Economics* **60**, 1–5. ID: 271683.

Guo, J.-F. and Ji, Q. (2013), 'How does market concern derived from the internet affect oil prices?', *Applied Energy* **112**, 1536–1543.
URL: *http://www.sciencedirect.com/science/article/pii/S0306261913002237*

Hallack, L. N. and Szklo, A. (2019), 'Assessing the exploratory potential in brazil by applying a creaming curve variant', *Energy Policy* **129**, 672–683.
URL: *https://www.sciencedirect.com/science/article/pii/S0301421519301521*

Hamilton, J. D. (1983), 'Oil and the macroeconomy since world war ii', *Journal of Political Economy* **91**(2), 228–248.

Hamilton, J. D. (1985), 'Historical causes of postwar oil shocks and recessions', *The Energy Journal* **6**(1), 97–116.
URL: *http://www.jstor.org/stable/41322100*

Hamilton, J. D. (2003), 'What is an oil shock?'. ID: 271689.

Hamilton, J. D. (2011), Historical oil shocks, Working Paper 16790, National Bureau of Economic Research.
URL: *http://www.nber.org/papers/w16790*

Hansen, B. E. (1997), 'Approximate asymptotic p values for structuras-change tests', *Journal of Business & Economic Statistics* **15**(1), 60–67.
URL: *https://doi.org/10.1080/07350015.1997.10524687*

Hassler, U. and Wolters, J. (2006), Autoregressive distributed lag models and cointegration, *in* O. Hübler and J. Frohn, eds, 'Modern Econometric Analysis: Surveys on Recent Developments', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 57–72.
URL: *https://doi.org/10.1007/3540326936_5*

Höök, M. and Aleklett, K. (2008), 'A decline rate study of norwegian oil production', *Energy Policy* **36**(11), 4262–4271. Transition towards Sustainable Energy Systems.
URL: *https://www.sciencedirect.com/science/article/pii/S0301421508003935*

Hubbert, M. K. (1962), Energy resources. a report to the committee on natural resources of the national academy of science-national research council., Technical Report 4, National Academy of Sciences-National Research Council.

Iledare, O. O. and Pulsipher, A. G. (1999), 'Sources of change in petroleum drilling productivity in onshore louisiana in the us, 1977–1994', *Energy Economics* **21**(3), 261–271.
URL: *https://www.sciencedirect.com/science/article/pii/S0140988399000092*

IMF (2019), 'World economic outlook update, july 2019: Still sluggish global growth'.
URL: *https://www.imf.org/en/Publications/WEO/Issues/2019/07/18/WEOupdateJuly2019*

International Energy Agency (2022), 'Oil 2021 - analysis and forecast to 2026', `https://www.iea.org/reports/oil-2021`.

Ji, Q. and Guo, J.-F. (2015), 'Oil price volatility and oil-related events: An internet concern study perspective', *Applied Energy* **137**, 256 – 264.
URL: *http://www.sciencedirect.com/science/article/pii/S0306261914010423*

Jin, X. (2019), 'The role of market expectations in commodity price dynamics: Evidence from oil data', *Journal of International Money and Finance* **90**, 1 –

18.
URL: *http://www.sciencedirect.com/science/article/pii/S0261560618305485*

Jo, S. (2014), 'The effects of oil price uncertainty on global real economic activity', *Journal of Money, Credit and Banking* **46**(6), 1113–1135.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/jmcb.12135*

Jurado, K., Ludvigson, S. C. and Ng, S. (2015), 'Measuring uncertainty', *American Economic Review* **105**(3), 1177–1216.
URL: *https://www.aeaweb.org/articles?id=10.1257/aer.20131193*

Kang, W. and Ratti, R. A. (2013), 'Structural oil price shocks and policy uncertainty', *Economic Modelling* **35**, 314 – 319.
URL: *http://www.sciencedirect.com/science/article/pii/S0264999313002964*

Kang, W., Ratti, R. A. and Vespignani, J. L. (2017), 'Oil price shocks and policy uncertainty: New evidence on the effects of us and non-us oil production', *Energy Economics* **66**, 536 – 546.
URL: *http://www.sciencedirect.com/science/article/pii/S0140988317300373*

Kaufman, G. M., Balcer, Y. and Kruyt, D. (1975), 'Probabilistic model of oil and gas discovery', *AAPG Studies in Geology* **1**, 113–142.

Kaufmann, R. (2011), 'The role of market fundamentals and speculation in recent price changes for crude oil', *Energy Policy* **39**(1), 105–115.
URL: *http://www.sciencedirect.com/science/article/pii/S0301421510007044*

Kaufmann, R. K. (1991), 'Oil production in the lower 48 states: Reconciling curve fitting and econometric models'. ID: 272271.

Kaufmann, R. K., Dees, S., Karadeloglou, P. and Sánchez, M. (2004), 'Does opec matter? an econometric analysis of oil prices', *The Energy Journal* **25**(4), 67–90.
URL: *http://www.jstor.org/stable/41323358*

Kellogg, R. (2011), 'Learning by drilling: Interfirm learning and relationship persistence in the texas oilpatch', *The Quarterly Journal of Economics* **126**(4), 1961–2004.

Kellogg, R. (2014), 'The effect of uncertainty on investment: Evidence from texas oil drilling', *American Economic Review* **104**(6), 1698–1734.
URL: *https://www.aeaweb.org/articles?id=10.1257/aer.104.6.1698*

Kemp, A. and Kasim, S. (2003), 'An econometric model of oil and gas exploration development and production in the uk continental shelf: A systems approach.', **24(2)**(The Energy Journal), 113–141.

Kemp, A. and Kasim, S. (2006), 'A regional model of oil and gas exploration in the ukcs', *Scottish Journal of Political Economy* **53**(2), 198–221.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9485.2006.00376.x*

Khalifa, A., Caporin, M. and Hammoudeh, S. (2017), 'The relationship between oil prices and rig counts: The importance of lags', *Energy Economics* **63**, 213 – 226.
URL: *http://www.sciencedirect.com/science/article/pii/S0140988317300257*

Kilian, L. (2008), 'Exogenous oil supply shocks: How big are they and how much do they matter for the u.s. economy?', *The review of economics and statistics* **90**(2), 216–240. doi: 10.1162/rest.90.2.216; 08.

Kilian, L. (2009), 'Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market', *The American Economic Review* **99**(3), 1053–1069.
URL: *http://www.jstor.org/stable/25592494*

Kilian, L. (2017), 'The impact of the fracking boom on arab oil producers', *The Energy Journal* **38**(6).

Kilian, L. and Murphy, D. P. (2014), 'The role of inventories and speculative trading in the global market for crude oil', *Journal of Applied Econometrics* **29**(3), 454–478.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.2322*

Klovland, J. T. (2004), Business cycles, commodity prices and shipping freight rates: Some evidence from the pre-wwi period, *in* 'Workshop on Market Performance and the Welfare Gains of Market Integration in History, European University Institute, Florence, Italy', Citeseer, pp. 1–4.

Kolodzeij, M. and Kaufmann, R. (2014), 'Oil demand shocks reconsidered: A cointegrated vector autoregression', *Energy Economics* **41**, 33 – 40.
URL: *http://www.sciencedirect.com/science/article/pii/S014098831300234X*

Kripfganz, S., Schneider, D. C. et al. (2018), ardl: Estimating autoregressive distributed lag and equilibrium correction models, *in* 'Proceedings of the 2018 London Stata Conference'.

Kupelian, B. and Jakeman, M. (2018), 'Global economy watch is us dollar debt in emerging markets a cause for concern?'.
URL: *https://www.pwc.com/gx/en/issues/economy/global-economy-watch/assets/pdfs/december-2018.pdf*

Laherrere, J. (2002), 'Forecasting future production from past discovery', *International Journal of Global Energy Issues* **18**(2-4), 218–238.

Lardic, S. and Mignon, V. (2008), 'Oil prices and economic activity: An asymmetric cointegration approach', *Energy Economics* **30**(3), 847–855.
URL: *https://www.sciencedirect.com/science/article/pii/S0140988306001265*

Leduc, S. and Liu, Z. (2016), 'Uncertainty shocks are aggregate demand shocks', *Journal of Monetary Economics* **82**, 20 – 35.
URL: *http://www.sciencedirect.com/science/article/pii/S0304393216300393*

Lee, J. and Strazicich, M. C. (2003), 'Minimum lagrange multiplier unit root test with two structural breaks', *The Review of Economics and Statistics* **85**(4), 1082–1089.
URL: *https://doi.org/10.1162/003465303772815961*

Lee, K., Ni, S. and Ratti, R. A. (1995), 'Oil shocks and the macroeconomy: the role of price variability', *Energy Journal* **16**(4), 39 – 56. Cited by: 474.
URL: *https://www.scopus.com/inward/record.uri?eid=2-s2.0-0029481013&doi=10.5547%2fISSN0195-6574-EJ-Vol16-No4-2&partnerID=40&md5=3e9a4e91bc039103cc2973ef4b52560a*

Lee, P.-J. and Wang, P. (1983), 'Probabilistic formulation of a method for the evaluation of petroleum resources', *Journal of the International Association for Mathematical Geology* **15**(1), 163–181.

Li, S., Zhang, H. and Yuan, D. (2019), 'Investor attention and crude oil prices: Evidence from nonlinear granger causality tests', *Energy Economics* **84**.
URL: *https://www.sciencedirect.com/science/article/pii/S0140988319302750*

Lillestøl, J. and Sinding-Larsen, R. (2017), 'Creaming and the likelihood of discovering additional giant petroleum fields', *Mathematical Geosciences* **49**(1), 67–83.

Lillestøl, J. and Sinding-Larsen, R. (2018), 'Creaming - and the depletion of resources: A bayesian data analysis', (No. 2017/16).
URL: *https://ssrn.com/abstract=3072338 or http://dx.doi.org/10.2139/ssrn.3072338*

Ludvigson, S. C., Ma, S. and Ng, S. (2015), Uncertainty and business cycles: Exogenous impulse or endogenous response?, Working Paper 21803, National Bureau of Economic Research.
URL: *http://www.nber.org/papers/w21803*

Lütkepohl, H. (1991), *Introduction to multiple time series analysis*, Springer - Verlag, Berlin.

MacDonald, D., Power, M. and Fuller, J. (1994), 'A new discovery process approach to forecasting hydrocarbon discoveries', *Resource and Energy Economics* **16**(2), 147–166.
URL: *https://www.sciencedirect.com/science/article/pii/0928765594900035*

Managi, S., Opaluch, J. J., Jin, D. and Grigalunas, T. A. (2006), 'Stochastic frontier analysis of total factor productivity in the offshore oil and gas industry', *Ecological Economics* **60**(1), 204 – 215.
URL: *http://www.sciencedirect.com/science/article/pii/S0921800905005756*

Mauritzen, J. (2017), 'The effect of oil prices on field production: Evidence from the norwegian continental shelf', *Oxford Bulletin of Economics and Statistics* **79**(1), 124–144.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/obes.12138*

McKinsey & Company (2021), 'Global oil supply-and-demand outlook to 2040'.
URL: *https://www.mckinsey.com/industries/oil-and-gas/our-insights/global-oil-supply-and-demand-outlook-to-2040*

Mohn, K. (2008), 'Efforts and efficiency in oil exploration: A vector error-correction approach', *The Energy Journal* **29**(4), 53 – 78.
URL: *http://www.iaee.org/en/publications/ejarticle.aspx?id=2277*

Mohn, K. and Osmundsen, P. (2008), 'Exploration economics in a regulated petroleum province: The case of the norwegian continental shelf', *Energy Economics* **30**(2), 303–320.
URL: *https://www.sciencedirect.com/science/article/pii/S0140988306001277*

Moore, A. (2017), 'Measuring economic uncertainty and its effects', *Economic Record* **93**(303), 550–575.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/1475-4932.12356*

Mork, K. A., Olsen, O. and Mysen, H. T. (1994), 'Macroeconomic responses to oil price increases and decreases in seven oecd countries', *The Energy Journal* **15**(4), 19–35.
URL: *http://www.jstor.org/stable/41322565*

Narayan, P. K. and Popp, S. (2010), 'A new unit root test with two structural breaks in level and slope at unknown time', *Journal of Applied Statistics* **37**(9), 1425–1438.
URL: *https://doi.org/10.1080/02664760903039883*

NEB (2018), 'Canada's energy future 2017 supplement: Conventional, tight, and shale oil production'.
URL: *https://www.cer-rec.gc.ca/en/data-analysis/canada-energy-future/2017-conventional/index.html*

Nehring, R. (1978), 'Giant oil fields and world oil resources'.

Norwegian Ministry of Petroleum and Energy and Norwegian Petroleum Directorate (2022), 'Exploration activity', `https://www.norskpetroleum.no/en/exploration/exploration-activity/`. Accessed: 2022-03-14.

Offshore Norge (2017), 'Norway's petroleum history', `https://offshorenorge.no/en/about-us/oljehistorien/`.

Osmundsen, P., Roll, K. H. and Tveterås, R. (2010), 'Exploration drilling productivity at the norwegian shelf', *Journal of Petroleum Science and Engineering* **73**(1), 122–128.
URL: *https://www.sciencedirect.com/science/article/pii/S0920410510001129*

Pesaran, M. H. (1990), 'An econometric analysis of exploration and extraction of oil in the u.k. continental shelf', *The Economic Journal* **100**(401), 367–390.

Pesaran, M. H. and Samiei, H. (1995), 'Forecasting ultimate resource recovery'. ID: 271676.

Pesaran, M. H., Shin, Y. and Smith, R. J. (2001), 'Bounds testing approaches to the analysis of level relationships', *Journal of Applied Econometrics* **16**(3), 289–326.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.616*

Pesaran, M. and Shin, Y. (1997), An autoregressive distributed lag modelling approach to cointegration analysis, Technical report, Faculty of Economics, University of Cambridge.

Pindyck, R. S. (1980), 'Uncertainty and exhaustible resource markets', *Journal of Political Economy* **88**(6), 1203–1225.

Ploberger, W. and Krämer, W. (1992), 'The cusum test with ols residuals', *Econometrica: Journal of the Econometric Society* pp. 271–285.

Power, M. and Jewkes, E. (1992), 'Simulating the discovery rate implications of technological innovation in oil and gas exploration', *Technological Forecasting and Social Change* **41**(2), 161–175.
URL: *https://www.sciencedirect.com/science/article/pii/004016259290061W*

Prest, B. C. (2018), 'Explanations for the 2014 oil price decline: Supply or demand?', *Energy Economics* **74**, 63 – 75.
URL: *http://www.sciencedirect.com/science/article/pii/S0140988318302020*

Qadan, M. and Nama, H. (2018), 'Investor sentiment and the price of oil', *Energy Economics* **69**, 42 – 58.
URL: *http://www.sciencedirect.com/science/article/pii/S0140988317303766*

Ramsey, J. B. (1980), 'The economics of oil exploration: A probability-of-ruin approach', *Energy Economics* **2**(1), 14–30.
URL: *https://www.sciencedirect.com/science/article/pii/0140988380900419*

Rao, R. D. (2000), 'An integrated modelling framework for exploration and extraction of petroleum resources1most of the work for this paper was

done during the author's stay at the indira gandhi institute of development research, mumbai, india.1', *Resources Policy* **26**(3), 133–143.
URL: *https://www.sciencedirect.com/science/article/pii/S0301420700000222*

Ratti, R. A. and Vespignani, J. L. (2015), 'Opec and non-opec oil production and the global economy', *Energy Economics* **50**, 364 – 378.
URL: *http://www.sciencedirect.com/science/article/pii/S0140988314003053*

Reynolds, D. B. (2002), 'Using non–time–series to determine supply elasticity: how far do prices change the hubbert curve?', *OPEC review* **26**(2), 147–167.

Ringlund, G. B., Rosendahl, K. E. and Skjerpen, T. (2008), 'Does oilrig activity react to oil price changes? an empirical investigation', *Energy Economics* **30**(2), 371 – 396.
URL: *http://www.sciencedirect.com/science/article/pii/S0140988307000771*

Rossi, B. and Sekhposyan, T. (2015), 'Macroeconomic uncertainty indices based on nowcast and forecast error distributions', *American Economic Review* **105**(5), 650–55.
URL: *https://www.aeaweb.org/articles?id=10.1257/aer.p20151124*

Sadorsky, P. (1999), 'Oil price shocks and stock market activity', *Energy Economics* **21**(5), 449 – 469.
URL: *http://www.sciencedirect.com/science/article/pii/S0140988399000201*

Salisu, A. A. and Fasanya, I. O. (2013), 'Modelling oil price volatility with structural breaks', *Energy Policy* **52**, 554–562. Special Section: Transition Pathways to a Low Carbon Economy.
URL: *https://www.sciencedirect.com/science/article/pii/S0301421512008609*

Scotti, C. (2016), 'Surprise and uncertainty indexes: Real-time aggregation of real-activity macro-surprises', *Journal of Monetary Economics* **82**, 1-19.
URL: *http://www.sciencedirect.com/science/article/pii/S0304393216300320*

Shi, S., Hurn, S. and Phillips, P. C. B. (2019), 'Causal Change Detection in Possibly Integrated Systems: Revisiting the Money–Income Relationship*', *Journal of Financial Econometrics* **18**(1), 158–180.
URL: *https://doi.org/10.1093/jjfinec/nbz004*

Shi, S., Phillips, P. C. B. and Hurn, S. (2018), 'Change detection and the causal impact of the yield curve', *Journal of Time Series Analysis* **39**(6), 966–987.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/jtsa.12427*

Shields, K. and Tran, T. D. (2019), 'Uncertainty in a disaggregate model: A data rich approach using google search queries', *CAMA Working Paper* (83).
URL: *https://ssrn.com/abstract=3492817*

Shin, Y., Yu, B. and Greenwood-Nimmo, M. (2013), 'Modelling asymmetric cointegration and dynamic multipliers in a nonlinear ardl framework', pp. 281–314.
URL: *https://ssrn.com/abstract̄1807745*

Sims, C. A. (1980), 'Comparison of Interwar and Postwar Business Cycles: Monetarism Reconsidered', *American Economic Review* **70**(2), 250–257.
URL: *https://ideas.repec.org/a/aea/aecrev/v70y1980i2p250-57.html*

Skjerpen, T., Storrøsten, H. B., Rosendahl, K. E. and Osmundsen, P. (2018), 'Modelling and forecasting rig rates on the norwegian continental shelf', *Resource and Energy Economics* **53**, 220–239.
URL: *https://www.sciencedirect.com/science/article/pii/S0928765517303044*

Smith, J. L. (1980), 'A probabilistic model of oil discovery', *The review of economics and statistics* **62**(4), 587–594.

Smith, J. L. and Paddock, J. L. (1984), 'Regional modelling of oil discovery and production'. ID: 271683.

Søren, J. (2003), *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford Scholarship Online.

URL: *https://oxford.universitypressscholarship.com/view/10.1093/0198774508.001.0001/acprof-9780198774501-chapter-2*

Sorrell, S., Miller, R., Bentley, R. and Speirs, J. (2010), 'Oil futures: A comparison of global supply forecasts', *Energy Policy* **38**(9), 4990–5003. cited By 75.

URL: *https://www.scopus.com/inward/record.uri?eid=2-s2.0-79955940443&doi=10.1016%2fj.enpol.2010.04.020&partnerID=40&md5=99aab16f66b1b90f3e78c15061c2420b*

Sorrell, S. and Speirs, J. (2014), 'Using growth curves to forecast regional resource recovery: approaches, analytics and consistency tests', *Phil. Trans. R. Soc. A* **372**(2006), 20120317.

StataCorp, L. (2017), 'Stata time-series reference manual release 15', *Stata Press College Station, TX* .

Swanson, N. R. (1998), 'Money and output viewed through a rolling window', *Journal of Monetary Economics* **41**(3), 455–474.

URL: *https://www.sciencedirect.com/science/article/pii/S0304393298000051*

Swierzbinski, J. (2013), *Economics of Exploration for and Production of Exhaustible Resources*, Vol. 2-3, Elsevier Inc.

URL: *https://www.scopus.com/inward/record.uri?eid=2-s2.0-85042695523&doi=10.1016%2fB978-0-12-375067-9.00162-5&partnerID=40&md5=f60420b5c5bf986076d9764df2d5c7aa*

Thomopoulos, N. T. (2013), *Essentials of Monte Carlo simulation: Statistical methods for building simulation models*, Springer Science & Business Media.

Toews, G. and Naumov, A. (2015), 'The relationship between oil price and costs in the oil industry', *The Energy Journal* **36**. Copyright - International Association for Energy Economics 2015; Last updated - 2020-06-10.

URL: *https://search.proquest.com/docview/2259547792?accountid=8155*

Tran, T. D., Vehbi, T. and Wong, B. (2019), 'Measuring uncertainty for new zealand using data-rich approach', *Australian Economic Review* **52**(3), 344–352.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/1467-8462.12339*

Vu, T. K. and Nakata, H. (2018), 'Oil price fluctuations and the small open economies of southeast asia: An analysis using vector autoregression with block exogeneity', *Journal of Asian Economics* **54**, 1–21.
URL: *https://www.sciencedirect.com/science/article/pii/S104900781630224X*

Yin, L. and Zhou, Y. (2016), 'What drives long-term oil market volatility? fundamentals versus speculation', *Economics* **10**(20), 1–27. Copyright - Copyright Universitaet Kiel Jul 22, 2016; Document feature - Tables; ; Graphs; Last updated - 2016-09-09.
URL: *https://search.proquest.com/docview/1811919058?accountid=8155*

Zivot, E. and Andrews, D. (1992), 'Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis', *Journal of Business & Economic Statistics* **10**(3), 251–70.
URL: *https://EconPapers.repec.org/RePEc:bes:jnlbes:v:10:y:1992:i:3:p:251-70*

Zou, C. (2013), Chapter 1 - introduction, *in* C. Zou, ed., 'Unconventional Petroleum Geology', Elsevier, Boston, pp. 1–26.
URL: *https://www.sciencedirect.com/science/article/pii/B9780123971623000013*