Emilie C. Løberg

What Determines the Level of Natural Gas Demand?

An Empirical Analysis Across and Within Aggregated Regions of the World

Master's thesis in Industrial Economics and Technology Management
Supervisor: Anne Neumann
July 2020

Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Economics and Management
Dept. of Industrial Economics and Technology Management

NTNU
Norwegian University of
Science and Technology

Emilie C. Løberg

What Determines the Level of Natural Gas Demand?

An Empirical Analysis Across and Within Aggregated Regions of the World

Master's thesis in Industrial Economics and Technology Management
Supervisor: Anne Neumann
July 2020

Norwegian University of Science and Technology
Faculty of Economics and Management
Dept. of Industrial Economics and Technology Management

NTNU
Norwegian University of
Science and Technology

Preface

This thesis was written by Emilie C. Løberg during the spring semester of 2020 at the Norwegian University of Science and Technology (NTNU). The thesis was written as an answer to the course TIØ4905, which is part of the master profile ”Managerial Economics and Operations Research” within the master's program ”Industrial Economics and Technology Management” at the Department of Industrial Economics and Technology Management.

I would like to thank supervisor Professor Anne Neumann at the Department of Industrial Economics and Technology Management for good advice and interesting discussions.

Trondheim, July 2020

Executive Summary

We live in a time where we see an increased focus regarding how to limit the environmental impact made on the planet. Efficiency measures targeting energy consumption and climate gas emissions can help combat these challenges. While a reduction in the level of consumption of non renewables, such as natural gas, over time can be part of the solution, it puts the industry sector in a challenging position. Natural gas plays a critical part in the production of several industrially manufactured products that are important to how multiple challenges are solved in todays society. Thus, a continued demand for natural gas is inevitable as long as no sufficient alternatives exist. In order to better plan for the future it is therefore important to understand what determines the demand of natural gas, and how the relationship between these determinants and the level of natural gas consumption, especially in the industry sector, may differ throughout the world.

When empirically analysing the consumption of natural gas, existing literature tend to focus on smaller geographical areas in which they analyse various levels of natural gas consumption **within** the chosen area. In this thesis the level of focus is lifted to revolve around three aggregated regions; OECD Europe, OECD Americas and OECD Asia & Oceania, which represent three of the larger natural gas markets throughout the world. For these regions, regression models including industry level consumption of natural gas and some of the most commonly used determinants of natural gas demand, GDP, natural gas prices and oil prices, are created and compared. The purpose of this is to highlight any possible differences in the relationships between the chosen level of natural gas consumption and determinants of demand **across** the three regions. Furthermore, in addition to the industry level consumption, an overall and sub-industrial level of natural gas consumption for OECD Europe is also analysed, in order to make the same kind of comparison of relationships, only **within** OECD Europe.

In this thesis the various cases are analysed using an Autoregressive Distributed Lag framework (ARDL) in which an Ordinary Least Squares estimator (OLS) is applied. The resulting coefficients sign and magnitude are used as measures to describe the aforementioned relationships. In addition, the various cases are tested for cointegration. If the presence of

cointegration is proven, an Error Correction reparameterization (EC) is also conducted, in order to split the relationships into long-run and short-run parts.

A majority of the results of the work done in this thesis show clear similarities in the signs and magnitudes, irrespective of region and level of natural gas consumption. There are some magnitudes that differ slightly from the rest. However, comparing the findings in this thesis with the findings of existing literature, these deviations are not unique. In fact, the findings in this thesis are for the most part in line with the findings of existing literature. What this tells us is when it comes to empirical analysis of natural gas consumption there exists similarities between regression models focusing on smaller geographical areas and regression models focusing on larger aggregated areas, in terms of the relationships they describe.

Sammendrag

Vi lever i en tid hvor vi ser et økt fokus på å begrense miljøpåvirkningen vi har på jorden. Effektiviseringstiltak rettet mot energiforbruk og utslipp av uønskede klimagasser kan bidra til å bekjempe disse utfordringene. En reduksjon i forbruket av ikke-fornybare energikilder, slik som naturgass, kan være en del av løsningen. Samtidig setter dette industrisektoren i en utfordrende posisjon. Naturgass spiller en kritisk rolle i produksjonen av flere industrielt tilvirkede produkter som er viktige med tanke på hvordan mange av samfunnets utfordringer løses idag. Som følge av dette vil en etterspørsel etter naturgass i fremtiden være uunngåelig, all den tid det ikke eksisterer fullgode alternativer. Dersom man skal planlegge best mulig for fremtiden er det derfor viktig å ha innsikt i hva som driver etterspørselen etter naturgass, og hvordan forholdet mellom disse driverne og nivået på naturgassforbruk, spesielt i industrisektoren, kan variere rundt om i verden.

Eksisterende litteratur tenderer til å fokusere på mindre geografiske områder når naturgassforbruk analyseres empirisk. Ofte analyseres ulike forbruksnivåer **innenfor** disse områdene. I denne oppgaven flyttes fokuset til å omhandle tre aggregerte regioner; OECD Europa, OECD Amerika og OECD Asia & Oseania, som her representerer tre av de største naturgassmarkedene i verden. For disse regionene etableres og sammenliknes regresjonsmodeller som omfatter industrielt naturgassforbruk, samt noen av de hyppigst brukte driverne for naturgassetterspørsel; BNP, naturgasspriser og oljepriser. Hensikten med dette er å belyse potensielle forskjeller i forholdene mellom det valgte forbruksnivået og dets drivere **på tverrs av** de tre regionene. I tillegg til å analysere industrielt naturgassforbruk analyseres også totalt naturgassforbruk og forbruket i en utvalgt underindustri for OECD Europa, slik at en tilsvarende sammenlikning kan gjøres også **innenfor** OECD Europa.

I denne oppgaven benyttes et Autoregressive Distributed Lag rammeverk (ARDL) sammen med en ordinær minste kvadraters estimator (OLS) som analyserammeverk. For å beskrive de nevnte forholdene benyttes fortegn og størrelse på koeffisientene i de ulike regresjonsmodellene. I tillegg testes det for kointegrasjon. Dersom en en regresjonsmodell viser tegn på kointegrasjon gjøres en Error Correction reparametrisering (EC). Dette gjøres for å dele de nevnte forholdene i en langsiktig og en kortsiktig del.

Majoriteten av resultatene i denne oppgaven viser tydelige likheter hva gjelder fortegn og størrelse på koeffisisentene. Dette er uavhengig av region og nivå på naturgassforbruk. Enkelte av koeffisientenes størrelser er noe annerledes sammenliknet med de andre. Samtidig er dette ikke et unikt funn dersom en sammenlikner funnene i denne oppgaven med funnene presentert i eksisterende litteratur. Denne sammenlikningen viser faktisk at funnene presentert i denne oppgaven i stor grad sammenfaller med funnene i eksisterende litteratur. Dette forteller oss at det med tanke på naturgassforbruk er likheter mellom empiriske analyser som fokusere på mindre geografiske områder og empiriske analyser som fokusere på større, aggregerte områder, dersom en ser på forholdene de beskriver.

Contents

1 Introduction **1**

2 Reference Framework **5**

2.1 Natural Gas . . . 5

2.2 OECD . . . 10

3 Research Methodology **11**

4 Literature Review **13**

4.1 Aspects of Interest . . . 13

4.2 Literature Search Strategy . . . 14

4.3 Results of the Literature Review . . . 15

5 Data - Choice of Variables **19**

6 Econometric Framework **23**

6.1 Time Series Data and Linear Regression . . . 23

6.2 Stationarity . . . 24

6.3 ARDL and Cointegration Analysis . . . 25

6.4 Use of ARDL . . . 29

6.5 Diagnostics - Testing of the Models . . . 31

6.6 Analysis of Results . . . 33

6.7 Summary of Econometric Framework . . . 34

7 Results **35**

7.1 Stationarity and Lag Specification . . . 35

7.2 ARDL-results . . . 38

7.3 Cointegration . . . 39

7.4 EC-results . . . 40

7.5 Diagnostics 41

8 Interpretation, Discussion, Limitations and Extensions 43

8.1 Analysis of Results 43

8.2 Discussion of Scientific Method 47

8.3 Generalizability and Possible Further Work 48

8.4 An Extension including Carbon Prices 49

9 Conclusion 51

A Appendix: Equations 53

B Appendix: Figures 55

B.1 Data 55

B.2 Quantile-Quantile Plots 61

B.3 Normal-Probability Plots 63

B.4 CUSUM Plots 65

C Appendix: Tables 67

C.1 Additional Tables - Model Results 67

C.2 Cointegration - Tables for the ARDL(1, 1, 1, 1)-models 79

C.3 Diagnostics - Tables for the ARDL(1, 1, 1, 1)-models 80

List of Figures

1 Share of Total Increase in Overall Primary Energy Consumption between 2016 and 2018 . . . 8

2 Change in Natural Gas Consumption between 2009 and 2018 . . . 9

3 Change in Fuel Consumption by Source between 2009 and 2018 . . . 9

4 Level and First Difference Plots of Natural Gas Consumption - Europe - Industry 55

5 Level and First Difference Plots of Natural Gas Consumption - Americas - Industry . . . 55

6 Level and First Difference Plots of Natural Gas Consumption - Asia & Oceania - Industry . . . 56

7 Level and First Difference Plots of Natural Gas Consumption - Europe - Overall 56

8 Level and First Difference Plots of Natural Gas Consumption - Europe - Sub-industry . . . 56

9 Level and First Difference Plots of GDP - OECD Europe . . . 57

10 Level and First Difference Plots of Natural Gas Price - OECD Europe . . . 57

11 Level and First Difference Plots of Oil Price - OECD Europe . . . 58

12 Level and First Difference Plots of GDP - OECD Americas . . . 58

13 Level and First Difference Plots of Natural Gas Price - OECD Americas . . . 59

14 Level and First Difference Plots of Oil Price - OECD Americas . . . 59

15 Level and First Difference Plots of GDP - OECD Asia & Oceania . . . 59

16 Level and First Difference Plots of Natural Gas Price - OECD Asia & Oceania 60

17 Level and First Difference Plots of Oil Price - OECD Asia & Oceania . . . 60

18 Quantile-Quantile Plots for Residuals - Europe - Industry . . . 61

19 Quantile-Quantile Plots for Residuals - Americas - Industry . . . 61

20 Quantile-Quantile Plots for Residuals - Asia & Oceania - Industry . . . 62

21 Quantile-Quantile Plots for Residuals - Europe - Overall . . . 62

22 Quantile-Quantile Plots for Residuals - Europe - Sub-industry . . . 62

23 Normal-Probability Plots for Residuals - Europe - Industry . . . 63

24 Normal-Probability Plots for Residuals - Americas - Industry . . . 63

25 Normal-Probability Plots for Residuals - Asia & Oceania - Industry 64
26 Normal-Probability Plots for Residuals - Europe - Overall 64
27 Normal-Probability Plots for Residuals - Europe - Sub-industry 64
28 CUSUM Plots - Europe - Industry 65
29 CUSUM Plots - Americas - Industry 65
30 CUSUM Plots - Asia & Oceania - Industry 66
31 CUSUM Plots - Europe - Overall 66
32 CUSUM Plots - Europe - Sub-industry 66

List of Tables

1 Cases Analysed in this Thesis . . . 3

2 List of OECD member countries . . . 10

3 Classification Key for Literature Review: Subsets of Determinants of Demand 15

4 Subsets of determinants of demand covered . . . 15

5 Comparison of Aspects of Interest in this Thesis and in the Existing Literature Reviewed . . . 16

6 Explanation of the Time Series Utilized . . . 21

7 ADF Tests for Dependent Variables in different cases . . . 36

8 ADF Tests for Independent Variables - OECD Europe . . . 36

9 ADF Tests for Independent Variables - OECD Americas . . . 36

10 ADF Tests for Independent Variables - OECD Asia & Oceania . . . 36

11 Fuller (1996) Critical Values for the Dickey & Fuller (1979) Unit Root Test (ADF Test) . . . 37

12 ARDL-results - Part 1 . . . 38

13 ARDL-results - Part 2 . . . 39

14 Test-statistics for the Pesaran et al. (2001) Bounds Test - Case 3 . . . 39

15 Kripfganz & Schneider (2018) Critical Values for the Pesaran et al. (2001) Bounds Test - Case 3 . . . 40

16 EC-results - Part 1 . . . 41

17 EC-results - Part 2 . . . 41

18 Diagnostic Tests for the ARDL-models . . . 42

19 Brown et al. (1975) CUSUM Test for the ARDL-models . . . 42

20 Critical Values for the Brown et al. (1975) CUSUM Test . . . 42

21 Extended ARDL-results - Europe - Industry - ARDL(1, 0, 0, 1) . . . 67

22 Extended ARDL-results - Americas - Industry - ARDL(1, 1, 0, 0) . . . 67

23 Extended ARDL-results - Asia & Oceania - Industry - ARDL(1, 0, 1, 0) . . . 67

24 Extended ARDL-results - Europe - Overall - ARDL(1, 0, 0, 1) . . . 68

25 Extended ARDL-results - Europe - Sub-industry - ARDL(1, 0, 0, 1) 68
26 Extended ARDL-results - Europe - Industry - ARDL(1, 1, 1, 1) 69
27 Extended ARDL-results - Americas - Industry - ARDL(1, 1, 1, 1) 69
28 Extended ARDL-results - Asia & Oceania - Industry - ARDL(1, 1, 1, 1) . . . 69
29 Extended ARDL-results - Europe - Overall - ARDL(1, 1, 1, 1) 70
30 Extended ARDL-results - Europe - Sub-industry - ARDL(1, 1, 1, 1) 70
31 Comparison of ARDL-results from different specifications - Europe - Industry 71
32 Comparison of ARDL-results from different specifications - Americas - Industry 71
33 Comparison of ARDL-results from different specifications - Asia & Oceania - Industry 71
34 Comparison of ARDL-results from different specifications - Europe - Overall . 72
35 Comparison of ARDL-results from different specifications - Europe - Sub-industry 72
36 Extended EC-results - Europe - Industry - ARDL(1, 0, 0, 1) 73
37 Extended EC-results - Americas - Industry - ARDL(1, 1, 0, 0) 73
38 Extended EC-results - Europe - Sub-industry - ARDL(1, 0, 0, 1) 74
39 Extended EC-results - Europe - Industry - ARDL(1, 1, 1, 1) 75
40 Extended EC-results - Americas - Industry - ARDL(1, 1, 1, 1) 75
41 Extended EC-results - Europe - Sub-industry - ARDL(1, 1, 1, 1) 76
42 Comparison of EC-results from different specifications - Europe - Industry . . 77
43 Comparison of EC-results from different specifications - Americas - Industry . 77
44 Comparison of EC-results from different specifications - Europe - Sub-industry 78
45 Test-statistics for the Pesaran et al. (2001) Bounds Test - Case 3 - ARDL(1, 1, 1, 1) 79
46 Kripfganz & Schneider (2018) Critical Values for the Pesaran et al. (2001) Bounds Test - Case 3 - ARDL(1, 1, 1, 1) 79
47 Diagnostic Tests for the ARDL-models - ARDL(1, 1, 1, 1) 80
48 Brown et al. (1975) CUSUM test for the ARDL-models - ARDL(1, 1, 1, 1) . 80

1 Introduction

We live in a time where we see an increased focus regarding how to limit the environmental impact made on the planet. Such limitations are attempted done through vast efficiency measures across countries and sectors throughout the world. These measures could for instance target the amount of energy consumed in order to run specific processes. They could also target the amount of unwanted emissions resulting from said processes. An example of an actor partaking in creating and facilitating possible solutions is the UN. They target these issues through their initiative called UNEP, or the UN Environment Program.

With respect to this increased focus on limiting environmental impact on the planet, the industry sector in particular is in a challenging position. Several of todays industrially manufactured products are fundamentally important to how a number of challenges are solved in society in present time. An example of this could be the use of concrete for critical infrastructure and housing. Another is the need for certain pharmaceutical products, chemicals and petrochemicals, where there are limited alternatives, or none at all.

While there is a large, global consensus that there is need for a decrease in the level of consumption of non renewables over time, as well as in the level of climate gas emissions, a full stop of extraction would put the production of multiple fundamentally important products at risk. As long as there exists no sufficient alternatives to these products there will be a continued demand in the years to come.

Multiple of these products are dependent on the usage of natural gas, for instance as fuel or as a raw material. As a result of this it is of value to better understand what role natural gas plays in the industry sector throughout the world. Such insight is of upmost value when trying to plan for the future.

This thesis seeks to empirically analyse the relationships between the level of natural gas consumption in the industry sector and a set of determinants of demands, using an Autoregressive Distributed Lag framework (ARDL). These relationships will be described through equations generated using empirical data used for estimating coefficients. The purpose of estimating these coefficients is to use them for inference, not for prediction and forecasting.

The aim of the thesis is to estimate a set of such equations for a group of regions representing some of the natural gas markets throughout the world. The purpose of this is to highlight possible differences and similarities in the regional relationships between the level of natural gas consumption and the selected determinants of demand.

The rationale for this choice is that due to technologically challenging infrastructure and transport possibilities, the volumes of natural gas transported across regions are limited. Thus, natural gas is likely to be consumed in areas close to the point of extraction. This opens up for the possibility that the relationships between the level of demand for natural gas and specific determinants of demand are regionally conditioned.

The selected regions are OECD Europe, OECD Americas and OECD Asia & Oceania. This selection is made due to two reasons. First, this thesis aims to compare the results from regions that are separated by a significant geographical distance. This applies for the regions suggested. The second reason is that the International Energy Agency have chosen to group the OECD countries in these regions in their Natural Gas Information reports (IEA, 2018). As a result of this, high quality, annual data on natural gas consumption for these countries are available for several years back.

This aggregation of larger areas stands in contrast to a lot of previous work done with respect to investigating the relationship between the level of natural gas consumption and determinants of demand. A literature review conducted for a project report done the fall semester of 2019 found that similar studies tend to focus on smaller geographical areas, rather than larger areas. Thus, by selecting larger areas, this thesis seeks to add new insight to the existing literature. This literature review will be presented further in Section 4.

In addition to the geographical grouping, a comparison will also be made with regards to possible differences and similarities across different levels of natural gas consumption within OECD Europe. In this part, overall, industrial and sub-industrial consumption of natural gas will be compared. The sub-industry of choice is the chemicals and petrochemicals industry. The format of comparison will be the same as for the regional comparison of industrial consumption of natural gas. Comparing across different levels of consumption is an approach that has been applied in existing literature, however not for this specific region in the way proposed in this thesis.

Based on the goals presented above, two primary research questions for this thesis are formulated:

1. **Are there any differences in the relationship between the level of industrial natural gas consumption and a set of the most commonly used determinants of demand, across different regions throughout the world?**
2. **Are there any differences in the relationship between various levels of natural gas consumption and a set of the most commonly used determinants of demand, within the same region, more specifically OECD Europe?**

These will be attempted answered through comparing the following cases, such as described above:

Part 1		
OECD Europe Industry	**OECD Americas Industry**	**OECD Asia & Oceania Industry**
Part 2		
OECD Europe Industry	**OECD Europe Overall**	**OECD Europe Sub-industry**

Table 1: Cases Analysed in this Thesis

From here on out the specific cases presented in Table 1 will be referred to without the use of ”OECD” for simplicity. ”OECD” will still be applied when the regions themselves are mentioned.

A common approach used by economists to measure and compare such relationships is the estimation of various elasticities of demand. This aspect will be touched upon briefly in Sections 4 and 8.

Furthermore, in addition to a comparison of the various results, the research questions will be discussed with respect to the problems raised above, revolving possible efficiency measures linked to both reducing the amount of non renewables used and the amount of climate gas emissions. This discussion will, among other things, cover how ”new” determinants of demand may have an impact in the future, and lead to disruptions in already established natural gas markets. Due to short time horizons regarding data availability, it is difficult to include these specific determinants of demand in the regression models suggested. They may however have a significant impact in the future, which will be elaborated.

The thesis consists of nine sections including this introduction. In Section 2 a broader framework relevant for this thesis is presented. Section 3 contains a presentation of the methodological approach chosen with respect to the research conducted. In Section 4 a review of a selection of existing literature within the chosen field of research is presented. In Section 5 a description of the data used in this thesis is given. Section 6 contains a theoretical description of the econometric framework applied in this thesis. In addition model specification for the various cases analysed are presented. Section 6 also presents aspects related to diagnostic testing and interpretation of results. In Section 7 the results from the empirical analysis as presented in Section 6 are presented. Section 8 contains an analysis of the results as well as a discussion of the scientific method utilised. Furthermore some possible extensions to the regression models presented in this thesis are proposed. Last, a conclusion is presented in Section 9.

2 Reference Framework

In this section a broader framework relevant for this thesis is presented. The first part focuses mainly on natural gas itself and the industrial use of it. This part is largely based on a similar reference framework made for the project report mentioned in Section 1. The second part gives a brief introduction to OECD.

2.1 Natural Gas

Natural gas is a resource with many applications, both as a fuel and as a raw material. Similar to oil it is found in reservoirs beneath the earths surface, and is a result of sediments and deposits being kept under high pressure over a long time. It is the third most consumed primary energy source worldwide, accounting for 24 % of the overall primary energy consumption as of 2018. Only oil and coal beats it out with shares of 34 % and 27 % respectively (BP, 2019).

In order for natural gas to be made available for various consumers, it has to be transported to the necessary locations. This transportation can be done either compressed, through pipelines, or liquefied, through shipping of liquefied natural gas (LNG) (Raus, 2014). These two transportation means cover short to medium distances and longer distances, respectively.

What both mediums have in common is that they are very capital-intensive (Raus, 2014). This is due to strict requirements being put on the the levels of pressure and temperature of the gas in both cases. Thus, the gas has to be processed accordingly, and the material surrounding it has to withstand these requirements. These aspects are significantly more complicated than the general transportation infrastructure for coal and oil. Thus, the cost of transportation is much higher for natural gas than it is for coal and oil.

Regarding natural gas pricing, the IEA identifies three distinct price regions throughout the world. These are Europe, North America and Asia-Pacific. Each of these regions have their own way of determining natural gas prices. Alongside these regions, the IEA also recognises four additional regions when talking of regional consumption and production of natural gas

(IEA, 2019a). These are South & Central America, Africa, the Middle East and the Commonwealth of Independent States; CIS.

In the North American market the mechanism for determining prices of natural gas is the Henry Hub. This mechanism mainly takes changes in supply and demand into account when determining the price.

In the European market prices were for a long time determined based on long term contracts. In an attempt of liberalising the natural gas market, structural changes has been proposed and implemented during the past decades. Among other measures many actors have pursued vertical unbundling. Over time the goal has been to increase competition between companies, and building down market barriers.

In the Asia-Pacific market prices have also mainly been determined based on long term contracts. These have had a tendency to be indexed to crude oil prices (Zaretskaya & Bradley, 2015). Although a majority of the pricing is performed based on long term contracts, there has been an increase of short term contracts the past couple of years.

One of the main reasons why long term contracts have been a preferred pricing mechanism is due to how it facilitates predictability and security of supply. Long term contracts may also help reduce the exercise of market power. The risk of this is higher with short term contracts. The idea has thus been that both consumers and producers benefit more from long term contracts.

The price level of natural gas within the various regions have differed significantly over the years. This can be explained due to the fact that a majority of the natural gas produced within the various areas are consumed in the same areas (BP, 2019). The differences in the level of regional consumption and production is usually covered through import and export, but the magnitudes of these volumes vary across the different regions.

North America, Europe and Asia-Pacific are net importers, whereas the four remaining regions are net-exporters to varying degrees. The volumes that are imported and exported through both North America and South & Central America almost cancel each other out, meaning their annual levels of production and consumption are almost the same. However, the volumes in North America are much larger than the volumes in South & Central America, both in terms of what is produced and consumed, and what is imported and exported. The remaining five regions all seem to have a clearer profile with respect to this; either having a larger exportation or importation.

Pipelines are, and have been, the dominating means of transportation measured in volume transported. However, Asia Pacific and the Middle East have seen a significant increase in the volumes of LNG exported over the past decade, even beating out the volumes transported by pipeline. Also Europe and Africa now see significant volumes being transported by LNG. Compared to both level of production and consumption, however, the volumes are rather small. Thus, their potential effect on the prices of natural gas within different areas are limited. As a result, there are few indications that the prices across regions are converging, as of 2018 (BP, 2019).

However, based on the current situation, it is of interest to note that the LNG market is expected to grow. The many U.S liquefaction projects that are expected to be finalised in the coming years, as well as other projects, are building up under these expectations (Zaretskaya & Bradley, 2015). Depending on how the growth develops, and how the contracts regarding pricing associated with these projects turn out, they may have a significant impact on natural gas prices world wide.

Something that is currently affecting the variations in price level of natural gas, is the price of substitutes (EIA, n.d.). In several of the applications for natural gas, there is a competition for market share with other raw materials. Price variations for the different substitutes may as a result have a direct impact on the price of natural gas in the different regions.

An aspect of interest with respect to this is the relationship between natural gas and crude oil. As fuels, one would assume that the two are substitutes, and that there thus should be a clear relationship between the pricing of the two. However, the respective pricing mechanisms does not necessarily reflect this (Ramberg & Parsons, 2012). There is an ongoing debate on this relationship, but it will not be elaborated further in this thesis.

In the past decade (2009-2018) the level of consumption of natural gas has had the highest increase with 31 % compared to other primary energy sources such as oil and coal who have seen increases of close to 15 % and 10 %, respectively. Between 2016 and 2018, close to 40 % of the overall consumption increase was covered by natural gas. This is illustrated in Figure 1. Still, it is important to note that the changes in level of consumption differs significantly across the various regions across the world (BP, 2019).

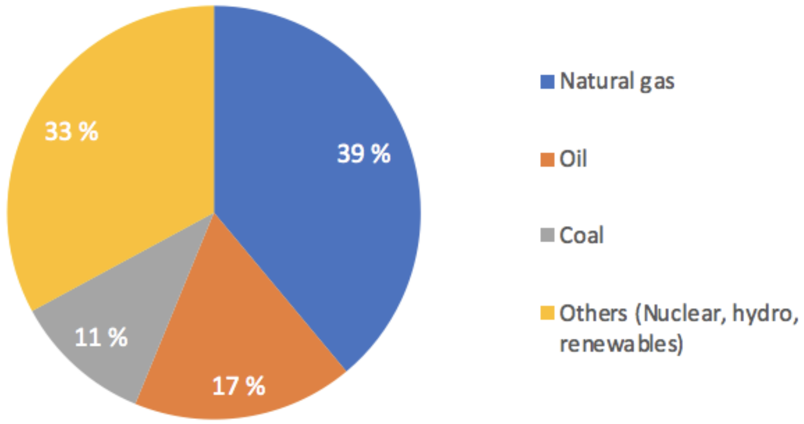


Figure 1: Share of Total Increase in Overall Primary Energy Consumption between 2016 and 2018. Source: BP (2019)

Between 2009 and 2018, North America and Asia-Pacific have seen a significant increase in level of natural gas consumption. The same applies to close to all of the other regions mentioned. Europe represents the only exception, where there has been a slight decrease. This is illustrated in Figure 2. In the same period of time, the level of consumption of both oil and coal has decreased in Europe, as illustrated in Figure 3. At the opposite end of the spectrum Asia-Pacific has had the biggest increases in level of consumption for all three resources in the same time period (BP, 2019).

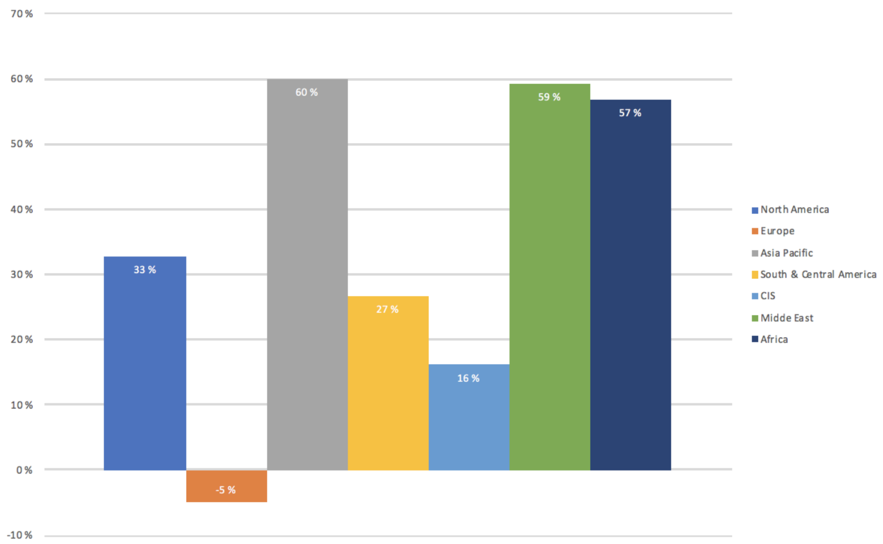


Figure 2: Change in Natural Gas Consumption between 2009 and 2018. Source: BP (2019)

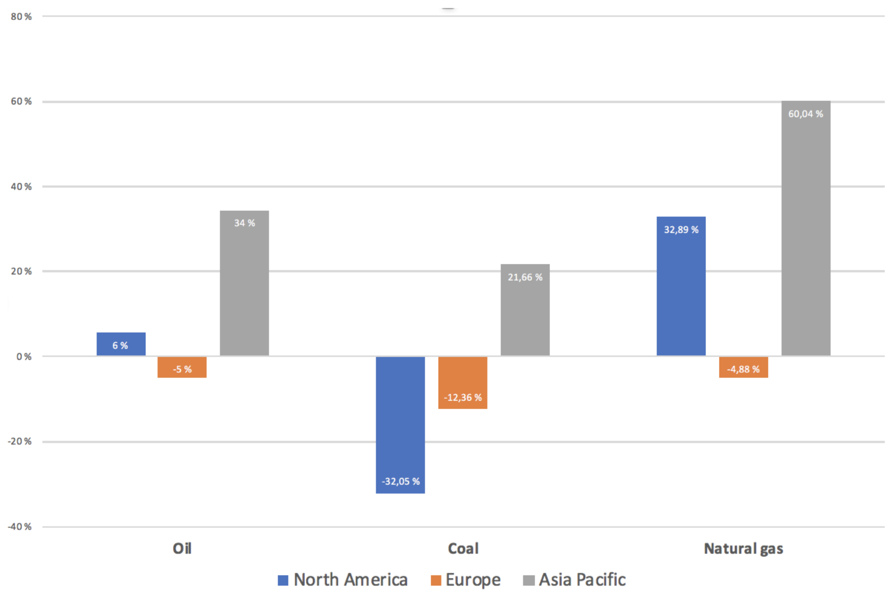


Figure 3: Change in Fuel Consumption by Source between 2009 and 2018. Source: BP (2019)

Among the contributors towards the overall demand for natural gas is the industry sector. In 2017, 40 % of the overall final energy consumption in the world stemmed from energy and industrial processes associated with the industry sector. A third of this was used for high-temperature heating (IEA, 2019b).

Alongside this, several of the various sub-industries not only contribute to a large consumption of primary energy sources. In the process, they also contribute to large amounts of CO_2 emissions. Among the various sub-industries the iron and steel, cement and chemical industries are some of the sources for large quantities of emissions. This is partly a result of process related emissions, often due to processes requiring high-temperature heating or from chemical reactions (IEA, 2019b).

While these emissions are undesired, they are hard to avoid as there are currently few alternatives to the application of primary energy sources in the industry sector. There are however initiatives looking into efficiency measures both with respect to the usage of various primary energy sources, but also with respect to mid- and end-of-process capturing of CO_2 emissions.

2.2 OECD

The OECD, or the Organisation for Economic Co-operation and Development, is a global forum working to create policies related to some of the major issues throughout the world. The purpose of the development of these policies is to help improve peoples lives over time. The OECD works on issues both on a regional, national and local level. Through their work they among other things supply large amounts of data, which can be utilised for analysis.

Currently there are 37 member countries from all over the world. The member countries are listed in Table 2 with year of accession for the respective countries listed in the parentheses. In addition, the OECD collaborates with a group of key partner countries.

OECD Europe			**OECD Americas**	**OECD Asia & Oceania**
Austria (1961)	Norway (1961)	Czech Republic (1995)	Canada (1961)	Israel (1962)
Belgium (1961)	Portugal (1961)	Poland (1996)	United States (1961)	Japan (1964)
Denmark (1961)	Spain (1961)	Hungary (1996)	Mexico (1994)	Australia (1971)
France (1961)	Sweden (1961)	Slovak Republic (2000)	Chile (2010)	New Zealand (1973)
Germany (1961)	Switzerland (1961)	Ireland (2010)		South Korea (1996)
Greece (1961)	Turkey (1961)	Slovenia (2010)		
Iceland (1961)	United Kingdom (1961)	Estonia (2010)		
Luxembourg (1961)	Italy (1962)	Latvia (2016)		
Netherlands (1961)	Finland (1969)	Lithuania (2018)		

Table 2: List of OECD member countries

3 Research Methodology

The following section contains a general presentation of the methodological approach chosen with respect to the research conducted in this thesis. The specific data and econometric framework utilized will be presented in Sections 5 and 6. Also, the search strategy applied for the literature review will be presented in Section 4.

In order to answer the research questions raised in Section 1 using an econometric method is an appropriate approach. As mentioned, this thesis seeks to evaluate the relationships between the level of natural gas consumption and a set of the most commonly used determinants of demand. Thus, a quantitative method combining economic theory and statistical models used for inference may help showcase these aspects. Similar research also tend to favor quantitative methods, so the approach chosen in this thesis would be in line with existing research on that matter.

The individual steps conducted for this thesis are presented briefly below:

Step 1 Literature Review:

- What are the most commonly used determinants of demand when empirically analysing the level of natural gas consumption?
- Which geographical areas are covered?
- Which levels of consumption are covered?
- Is there a focus on comparing results **across** areas?
- Is there a focus on comparing results **within** areas?
- What econometric methods are applied?

Step 2 Collection and evaluation of data:

1. Based on the findings of the literature review with respect to commonly used independent variables, collect relevant data.
2. Evaluate the quality of the available data, and decide on which data sets to use further.

Step 3 Modelling:

1. Construct appropriate regression models based on the results of the previous step.
2. Model the cases using an ARDL-framework
3. Test for the presence of cointegration in the various regression models.
4. If cointegration is present, conduct an Error Correction-reparameterization (EC-reparameterization).

Step 4 Analyse the results and discuss them with respect to the research questions raised in Section 1

4 Literature Review

The following section contains a review of a selection of existing literature that covers aspects related to demand of natural gas.

First, the main focus areas of the review is described. Second, the strategy utilized when searching for relevant literature is explained. Third, follows a review in which the selected literature are classified based on a selection of criteria chosen with respect to the specific aspects of interest in this thesis.

This section is largely based on a literature review conducted for the aforementioned project report. This means that the search strategy and the papers reviewed are exactly the same. However, some adjustments have been made with respect to the aspects of interest. Based on the classifications made, brief comments will be made on how this thesis separates itself from the literature reviewed.

4.1 Aspects of Interest

In the review the following aspects will be highlighted:

Subsets of determinants of demand covered

A classification of the independent variables, also known as determinants of demand, used in the various papers will be conducted. A key for this classification is presented in Table 3. The purpose of this being in focus is to use the results as decision support when deciding on which independent variables to include in the regression models for this thesis.

Choice of geographical area and levels of consumption covered

The various geographical areas and levels of consumption covered will be presented. This is of interest as there are different approaches possible when analysing the level of consumption of natural gas. This comes down to the difference in analysing aggregate areas and aggregate consumption, compared to a disaggregate approach, focusing on smaller areas and disaggregate consumption.

Choice of aspects for comparison

If the paper of interest conducts some kind of comparison in relation to their results, this will be highlighted. For this to be possible, it is a requirement that multiple cases are analysed in the specific papers. This is of interest due to it being a large focus in this thesis.

Method of choice

The econometric methods chosen will be briefly highlighted. This will cover differences in how the various papers are modelled with respect to the framework and estimators chosen.

4.2 Literature Search Strategy

The search for relevant literature was conducted in three stages.

In stage one, Google Scholar and Oria was utilised, searching for publications using the phrase ”determinants of demand of natural gas”. Combinations of the phrases ”determinants of demand” and ”natural gas demand” were also used. This resulted in a list of publications.

Stage two consisted of evaluating the suggested publications. In this evaluation, the aim was to discover citations of older papers that could be of interest to the literature review. These older papers were added to the list of relevant literature if they were found interesting.

In stage three a forward citation search was conducted using Google Scholar and Oria. This was done in order to discover more recent publications in which papers already on the list of relevant literature were cited.

For all papers added in either stage two or three, stage two and three were repeated. The purpose of this was to ensure that all papers on the list had been evaluated similarly.

4.3 Results of the Literature Review

Based on the key presented in Table 3 a classification of the various papers reviewed is presented in Table 4. The classifications will not be elaborated, as this is beyond the scope of this review. However, a brief discussion based on this classification is made in Section 5, where the independent variables of choice for this thesis are presented.

Subsets of determinants of demand
Subset 1: Price of good/service
Subset 2: Prices of related goods/services
Subset 3: Tastes/preferences of consumers
Subset 4: Income of consumer
Subset 5: Expectations of the consumers
Subset 6: Number of consumers

Table 3: Classification Key for Literature Review: Subsets of Determinants of Demand

Paper	**Subsets of determinants of demand covered**
Gautam & Paudel (2018)	1, 2, 3, 4
Zhang et al. (2018)	1, 2, 4
Burke & Yang (2016)	1, 3, 4, 6
Harold et al. (2015)	3, 6
Ackah (2014)	1, 3, 4, 6
Dilaver et al. (2014)	1, 3, 4
Yu et al. (2014)	1, 2, 3, 4, 6
Andersen et al. (2011)	1, 2, 4
Wadud et al. (2011)	1, 4, 6
Erdogdu (2010)	1, 4

Table 4: Subsets of determinants of demand covered

The results with respect to the remaining aspects of interest presented above are presented in Table 5. In this table, the specifications chosen for this thesis are also presented. Any possible similarities with the reviewed literature with respect to this are highlighted in the table.

Paper	Level of consumption	Area(s)	Comparisons	Econometric framework
This thesis	***Overall*** ***Industry*** ***Sub-industry***	***Multiple regions***	***Within area*** ***Across areas***	***ARDL-specification*** ***Ordinary Least Squares estimator***
Gautam & Paudel (2018)	Residential **Industry** Commercial	Sub-country	Across estimators **Within area**	Panel data model **ARDL-specification** Dynamic Fixed Effects estimator Mean Group estimator Pooled Mean Group estimator Common Correlated Effect Mean estimator Augmented Mean Group estimator
Zhang et al. (2018)	Power generation Residential **Industry** Service Transportation	Country	**Within area**	**ARDL-specification**
Burke & Yang (2016)	**Overall** Residential **Industry**	**Multiple countries aggregated**	Across approaches Across estimators	Panel data model Single equation approach Instrumental Variable approach Between estimator Pooled Ordinary Least Squares model Fixed Effects estimator
Harold et al. (2015)	Residential	Country	Other	Panel data model Random Effects estimator
Ackah (2014)	**Overall** Residential **Industry**	Country	**Within area**	Structural Time Series Model **ARDL-specification**
Dilaver et al. (2014)	**Overall**	**Region**	-	Structural Time Series Model **ARDL-specification**
Yu et al. (2014)	Residential	Sub-country	-	Panel data model Feasible General Least Squares estimator
Andersen et al. (2011)	**Sub-industries**	Multiple countries	**Within areas** **Across areas**	Panel data model Shrinkage estimator
Wadud et al. (2011)	**Overall** Power generation Residential **Industry** Fertilizer	Country	**Within area**	Partial adjustment specification
Erdogdu (2010)	Power generation Residential **Industry**	Country	**Within area**	Partial adjustment specification

Table 5: Comparison of Aspects of Interest in this Thesis and in the Existing Literature Reviewed

Looking at the findings presented in Table 5 it is evident that there are some clear similarities as well as differences with respect to aspects of interest in this thesis compared with the literature reviewed.

A majority of the papers analyse several levels of natural gas consumption. Furthermore only one geographical area tend to be preferred, and the results of the analysed levels of consumption are often compared within this area. Only Burke & Yang (2016) and Dilaver et al. (2014) choose to use aggregated regions. However only one region is chosen in each case, which stands in contrast to the approach chosen in this thesis, where multiple regions are covered.

What is worth noting is that Dilaver et al. (2014) also analyse OECD Europe. However, they make no comparisons. This is due to the fact that the purpose of their work is to estimate an Underlying Energy Demand Trend (UEDT). Thus, while the area of interest is the same, there is a difference in the purpose of the work done.

When it comes to the choice of econometric framework, there is no clear preference across the papers reviewed. Multiple specifications and estimators are used. In the cases of Gautam & Paudel (2018) and Burke & Yang (2016), where multiple estimators are used, the results across the chosen estimators are compared. This stands in contrast to the approach in this thesis where only one estimator, the Ordinary Least Squares estimator, is used. Furthermore, in five out of the ten papers reviewed analyses across the various cases are conducted using a panel data model. This also stands in contrast to the approach of this thesis, where all cases are analysed separately from one another.

The work conducted by Andersen et al. (2011) is the most similar to this thesis with respect to overall approach. They analyse several levels of natural gas consumption in multiple countries, utilising only one estimator. Furthermore, the results are compared both within a geographical area, as well as across the various areas. What separates the work conducted by Andersen et al. (2011) from this thesis is the actual choice of levels of consumption and geographical areas along with econometric framework.

Based on these findings the choice of variables used in this thesis will be presented in Section 5.

In addition to the aspects highlighted in Table 5 it is worth mentioning that a majority of the papers reviewed focus largely on various elasticities, such as income and own price elasticities of demand. While this, as mentioned in Section 1, is a commonly used approach when measuring and comparing the relationships between dependent and independent variables, it is not the main focus in this thesis. A brief comparison, will however be made between the results presented in this thesis and the results from the reviewed literature. This comparison will be made with respect to the sign and magnitude of the estimated coefficients in order to assess whether or not the results presented in this thesis are realistic and in line with both economic theory and the existing literature reviewed. This will be presented in Section 8.

5 Data - Choice of Variables

The following section gives a description of the data used in this thesis. A brief argumentation for the choices will be presented. This argumentation will be elaborated more in Section 6.3.2.

As independent variables in this thesis, five different forms of natural gas consumption data have been chosen. The OECD database provides nuanced, annual data dating back as early as 1971. For this thesis the chosen consumption data includes overall consumption, consumption hailing from the overall industry sector, and consumption from a specific sub-industry, the chemicals and petrochemicals industry. As mentioned in Section 1 several cases will be analysed in this thesis. First, the overall industrial natural gas consumption for OECD Europe, OECD Americas and OECD Asia & Oceania will be analysed separately and compared. Second, the three levels of consumption presented above for OECD Europe will be analysed separately and compared.

The independent variables chosen for the various regression models will all be the same ones, only regionally specific. The choice of variables is based on the results of the literature review presented in Section 4. One of the aims of the review was to investigate what were the most commonly used 'determinants of demand' when empirically analysing the demand for natural gas.

Based on the findings presented in Table 4 it is apparent that some subsets of determinants of demand are more commonly used than others. The various papers reviewed have in common that they include variations of three specific independent variables in their regression models. First, they tend to include a variable that relates to the income of the geographical area of interest, such as GDP or GDP per capita (Subset 4). Sometimes these kinds of variables are used in combination with a variable representing the size of the population in the area (Subset 6). Second, own prices are normally included (Subset 1). Last, the price of one or more substitutes tend to be included (Subset 2). In addition to these, the various papers may include variables that are inherently relevant to their specific cases. An example of this can be the inclusion of a climate based variable when analysing residential demand for natural gas.

These findings make out the foundation of the regression models in this thesis. All cases will be analysed using a base formulation including GDP, average natural gas price and average oil price as independent variables. The GDP and natural gas prices are both specific to the region of interest, while the oil price utilised is common across all cases analysed.

While most of the papers reviewed use real data, in this thesis the nominal versions of the three independent variables are utilised. This is due to the added complexity of converting the aggregated GDP of an entire region from nominal to real. This is beyond the scope of this thesis, and thus all independent variables have been kept as nominal data.

The various variables used are presented in Table 6. The natural gas and oil prices acquired are used as proxies for regional natural gas prices and a global oil price.

What is worth mentioning is that there are fewer available data points for the Japan - average natural gas price, used as a proxy for the natural gas price in OECD Asia & Oceania. As a result, when analysing the case representing OECD Asia & Oceania, the remaining variables are adjusted accordingly. A consequence of this is that this single case has a different number of observations than the four other cases. A consequence of this is that a direct comparison between the various models should not be performed. This is due to the fact that regression models should consist of the same amount of data points for a comparison to make proper sense. A solution to this could be to adjust the remaining four cases accordingly in order for the results to be directly comparable. A problem with this solution is that it reduces the amount of available data points for the other cases significantly, which is not ideal, especially since the sample sizes are rather small already, due to the frequency. A second option could be to drop the case representing OECD Asia & Oceania all together, but then a dimension to the thesis is lost.

The chosen approach in this thesis is to keep all cases as they are, and not adjust the remaining four cases for the smaller amount of available data points for the case representing OECD Asia & Oceania. When comparing the cases this will, however, have to be kept in mind, as this represents a flaw in the method for this specific case.

Time Series [Unit]	**Description**	**Data Source**	**Frequency**	**N#**
Consumption of Natural Gas - Europe - Industry [million m^3]	Aggregate industry-level consumption of natural gas for OECD Europe	OECD iLibrary - IEA Natural Gas Information Statistics	Annual	47
Consumption of Natural Gas - Americas - Industry [million m^3]	Aggregate industry-level consumption of natural gas for OECD Americas	OECD iLibrary - IEA Natural Gas Information Statistics	Annual	47
Consumption of Natural Gas - Asia & Oceania - Industry [million m^3]	Aggregate industry-level consumption of natural gas for OECD Asia & Oceania	OECD iLibrary - IEA Natural Gas Information Statistics	Annual	41
Consumption of Natural Gas - Europe - Overall [million m^3]	Aggregate consumption of natural gas for OECD Europe	OECD iLibrary - IEA Natural Gas Information Statistics	Annual	47
Consumption of Natural Gas - Europe - Sub-industry [million m^3]	Aggregate consumption of natural gas for OECD Europe for the chemicals and petrochemicals industry	OECD iLibrary - IEA Natural Gas Information Statistics	Annual	47
GDP - OECD Europe [million USD]	Aggregate Gross Domestic Product for OECD Europe	OECD iLibrary	Annual	47
GDP - OECD Americas [million USD]	Aggregate Gross Domestic Product for OECD Americas	OECD iLibrary	Annual	47
GDP - OECD Asia & Oceania [million USD]	Aggregate Gross Domestic Product for OECD Asia & Oceania	OECD iLibrary	Annual	41
Natural Gas Price - Europe [USD/mmbtu]	Average natural gas price for Europe	World Bank	Annual	47
Natural Gas Price - US [USD/mmbtu]	Average natural gas price for the US	World Bank	Annual	47
Natural Gas Price - Japan [USD/mmbtu]	Average natural gas price for Japan	World Bank	Annual	41
Crude Oil, Average [USD/bbl]	Average crude oil price	World Bank	Annual	41/47

Table 6: Explanation of the Time Series Utilized

6 Econometric Framework

In this section the econometric framework utilised in this thesis is presented. First, a presentation of general econometric theory and methodology that is relevant to this thesis is given in Subections 6.1 - 6.3. Second, in Subsection 6.4 the regression model specifications of the various cases analysed in this thesis are presented. Third, in Subections 6.5 and 6.6 areas related to testing and interpretation of the results are presented. Last, a summary of the econometric framework is listed.

6.1 Time Series Data and Linear Regression

Time series data are data points represented in a list or vector. The list contains a specific type of observation, sampled at a given frequency and sorted in order of time of observation (Brooks, 2014). An example of such a list is $Y = [y_1, y_2, ..., y_T]$, where the list contains T observations, and y_t denotes the observation at time t.

Time series data can be used to gain better knowledge on the relationship between different variables. Linear regression is a commonly used technique to accomplish this. The most basic regression model, which contains only one independent variable, is presented in equation (6.1) (Woolridge, 2009).

$$y_t = \beta_0 + \beta_1 x_t + \nu_t \tag{6.1}$$

Here, y_t represents a dependent variable and x_t represents the single independent variable in this model. ν_t represents the error-term, which contains the residuals from the equation. All three are vectors of the same length. β_0 and β_1 represent the intercept and slope coefficients, and the goal is for these coefficients to be estimated. This is done by utilising an estimator.

The Ordinary Least Squares estimator (OLS) is one of the most commonly used estimators. The coefficients are estimated based on what best minimizes the square of the error terms. Under the Gauss-Markov Assumptions OLS estimators are viewed as the Best Linear

Unbiased Estimators. As long as the assumptions hold, the OLS estimators are viewed as consistent, unbiased and efficient (Woolridge, 2009). As a result of this, OLS is the estimator of choice in this thesis. In order to control that the regression models in this thesis are indeed consistent, unbiased and efficient, a set of tests based on the Gauss-Markov Assumptions will be performed.

6.2 Stationarity

In order to use a specific time series in a regression model, its statistical properties need to be investigated. This has to be done in order to determine if the time series is stationary. When a time series is stationary its mean, variance and autocovariance are all constant over time (Brooks, 2014).

If non-stationary time series are used in a regression model, one risks that the regression becomes spurious. This is a term used when the output of the regression, in the form of coefficients and the models goodness of fit, depict an existing relationship between a set of variables, even if this is not statistically true (Brooks, 2014). Thus, in order to ensure that the output of a regression is not deceptive, the stationarity of the included variables needs to be investigated.

When talking about the stationarity of a time series, the term order of integration, denoted by $I(d)$, is often used. Here, d represents the order of integration, and is the number of times the specific time series needs to be differenced in order for it to be stationary. If a time series does not need to be differenced in order to be stationary it is stationary at levels, which is denoted by $I(0)$ (Brooks, 2014).

A commonly used method in order to investigate the order of integration of a time series is the Dickey & Fuller (1979) unit root test, more commonly known as the Augmented Dickey Fuller test (ADF test). It tests for the presence of a unit root in the time series. A unit root represents the presence of a stochastic trend in the time series, and it indicates non-stationarity (Brooks, 2014). The null hypothesis of the test is that a unit root is present.

There are multiple benefits associated with the ADF test. First, it allows the researcher to specify the inclusion of drifts or trends in the time series, if that is the case for the time series under investigation. Second, it allows for the inclusion of lagged differences. These are both important nuances that have an impact on how the specific variable will be used in the regression model of choice. Thus, it is important that these aspects are reflected when a test of stationarity is performed.

In addition to the ADF test there exist multiple other tests that serve the same purpose. Examples of some of these tests are the Phillips-Perron test and the Dickey-Fuller GLS test. While, according to Brooks (2014), the various tests tend to give the same results, comparing results of stationarity tests is beyond the scope of this thesis. Thus, the ADF test is the only test that will be used.

Furthermore, visual inspections of the line plots of the level and first differences of the variables of interest will be performed. These types of plots can serve as support alongside the ADF test, in order to assess the order of integration of the various variables utilised in this thesis.

6.3 ARDL and Cointegration Analysis

When investigating the relationship between a set of variables, it is plausible to believe that the events of previous time periods may have an impact on the current and future time periods. When modelling using regression, this concept is introduced through lag terms, which are included as separate terms on the right-hand side of the equation. This changes the regression model from a static model to a dynamic one. When lags of the dependent variable are included in the model, the model is said to be autoregressive. When lags of independent variables are included, the model is said to contain distributed lags. A model containing both kinds of lags is called an Autoregressive Distributed Lag model (ARDL) (Brooks, 2014).

Pesaran & Shin (1999) introduced the use of ARDL-models for the purpose of cointegration analysis. The work of Pesaran et al. (2001), which is based on Pesaran & Shin (1999), is now among the most popular methods for cointegration analysis. Cointegration analysis is the research into long-run relationship between a set of variables (Brooks, 2014). If a long-run relationship is proven to exist, it is possible to divide the relationship between variables into a short-run and a long-run part. In traditional cointegration theory the various variables are required to be integrated of the same order in order for a long-run relationship to exist. However, the method developed by Pesaran et al. (2001) allows for long-run relationships to exist between variables that are integrated of different orders, as long as none of them are integrated of order 2 or higher. This makes the approach suggested by Pesaran et al. (2001) more flexible than comparable methods.

The method developed by Pesaran et al. (2001) incorporates a bounds test in which a regressions F-statistic and t-statistic are compared against a set of pre-calculated confidence intervals. The intervals, generated by Pesaran et al. (2001), represent boundaries for whether or not a long-run relationship exist between the dependent and the independent variables in a regression. Pesaran et al. (2001) calculated a variety of different intervals depending on both the confidence level of choice, and on five different classes of model specifications. The lower bounds of the intervals are calculated based on the case where all variables in the regression model are cointegrated of order 0. The upper bounds are calculated based on the case where all variables in the model are cointegrated of order 1.

Kripfganz & Schneider (2018) have also calculated similar kinds of confidence intervals with the purpose of them being utilised in the same type of bounds test. In contrast to Pesaran et al. (2001) however, where the bounds are near-asymptotic, Kripfganz & Schneider (2018) have calculated two sets, where one has bounds that are finite-sample and the other has bounds that are asymptotic. In this thesis, the two varieties laid forward by Kripfganz & Schneider (2018) will be used when testing for cointegration.

6.3.1 ARDL-specification

When talking about a specific ARDL-model, it is normal to denote the lag specification by writing ARDL(a, b, c, ...) or simply (a, b, c, ...). Here, a represents the amount of lag terms of the dependent variable, where as b and c represent the amount of lag terms of two separate independent variables. Based on the format of equation (6.1), a general model including one independent variable, x, denoted by ARDL(a, b), is presented in equation (6.2) (Kripfganz & Schneider, 2018).

$$y_t = \alpha_0 + \sum_{i=1}^{a} \alpha_i y_{t-i} + \sum_{j=0}^{b} \beta_j x_{t-j} + \nu_t \tag{6.2}$$

This equation can be reparameterized into an equation called the conditional error correction equation (EC). In the EC-equation the coefficients are split into long-run and short-run parts. The EC-parameterization of equation (6.2) is presented in equation (6.3) (Kripfganz & Schneider, 2018).

$$\Delta y_t = \alpha_0 + \sum_{i=1}^{a} \alpha_i^* \Delta y_{t-i} + \sum_{j=0}^{b} \beta_j^* \Delta x_{t-j} - \lambda \left(y_{t-1} - \Theta x_t \right) + \nu_t \tag{6.3}$$

$\lambda = 1 - \sum_{i=1}^{a} \alpha_i$ represents the coefficient for speed of adjustment and serves the purpose of adjusting the error of the previous time period. λ is required to be negative and significant in order for the model to converge towards equilibrium. Otherwise, the errors are not adjusted for (Kripfganz & Schneider, 2018). $\Theta = \frac{\sum_{j=0}^{b}}{\lambda}$ represents the long-run coefficient of the independent variable in this model. For a model containing several independent variables, separate terms representing these variables are included to Θ. α_i^* and β_j^* represent the short-run coefficients of the dependent and independent variables, respectively.

A requirement for the EC-formulation to make practical sense is that there is a proven long-run relationship in the ARDL-model. Otherwise, splitting the ARDL-coefficients into long-run and short-run parts will not serve any meaningful purpose. It is the EC-formulations of the various cases that will be used when testing for cointegration in this thesis. Coefficients from these formulations will, however, only be presented if a long-run relationship is proven for the specific cases.

6.3.2 Model Specification

Before starting to model it is important to decide on a model specification. Specifying a regression model includes the processes of selecting which variables to include, and also deciding on lag specifications for the included variables. The goal behind such decisions is to achieve a model that performs well. What performing well means can be interpreted in different ways. Often the most parsimonious model is desired. This involves penalising the use of excess 'resources', which in this case means lags. When it comes to lag specification, a variety of information criteria are often used as tools to assess this (Brooks, 2014).

Two of the most used information criteria are the Akaike's (1974) information criterion (AIC) and the Schwarz's (1978) Bayesian information criterion (SBIC). They differ in the way they penalize the use of 'resources', or lags. The AIC is more lenient with regards to penalising lags, and are viewed to give larger regression models that are efficient, but not necessarily consistent. The SBIC on the other hand penalises lags harder, thus giving models that are smaller and consistent, but not as efficient as the models based on the AIC (Brooks, 2014).

When analysing the resulting coefficients of a regression model it is important to consider the effects that the lag specification may have. In relation to this, a risk associated with a too large lag structure is that the individual coefficients lose their descriptive value, as the overall effect from a specific variable is separated across a group of lags. In the work with this thesis, regression models using both information criteria have been considered, but based on the aforementioned argument the SBIC is the information criterion of choice in this thesis.

As mentioned above, another aspect regarding model specification that is important to consider is the choice of independent variables. Including too many independent variables increases the risk of the model being overfit (Brooks, 2014). In contrast to this, including too few variables increases the risk that the chosen variables do not carry enough descriptive power associated with the dependent variable in the model. This can for instance be seen through a large error term in the model. With respect to these aspects it is important to be conscious as to why specific variables are included in the regression model.

As presented in Section 5 a set of standard variables have been selected as independent variables for this thesis. When working with the thesis various variables have been included and removed from this set in order to assess the impact the individual variables have on the specific models in this thesis. Alongside the evaluation of various possible lag-structures, this underlines the fact that the modelling approach chosen for this thesis can be defined as a general-to-specific modelling approach. A discussion of the general-to-specific modelling approach, as well as the description and discussion of the process behind selecting independent variables, is beyond the scope of this thesis, and will not be elaborated further, with one exception which will be discussed in Section 8.

Another important aspect regarding the choice of independent variables is the requirement that these variables are all exogenous, meaning that their values are determined outside of the regression equation. If this is not the case, the requirements for the use of the OLS estimator do not hold (Brooks, 2014). Applying the ARDL-framework suggested above works around this specific requirement and results in valid coefficients even in the presence of endogenous variables. This was shown through the work of Pesaran & Shin (1995).

6.4 Use of ARDL

The following subsection gives a description of the regression models used in this thesis, and the rationale behind their specifications.

6.4.1 Double Log Models

All models in this thesis are double log models, meaning that all the values used for the various variables are the natural log of the original values extracted from the various sources presented in Section 5. This will not be written out in clear text in the following equations due to simplicity, but is important nonetheless.

The use of the double log specification allows for the interpretation of the resulting coefficients as elasticities of demand. In addition using the log-transformation of the variables leads to a reduction of the standard errors, which is an added bonus.

6.4.2 General Equations

Based on equation (6.2), equation (6.4) represents the general ARDL(a, b, c, d)-model that is utilised in this thesis:

$$y_{n,m,t} = \alpha_0 + \sum_{i=1}^{a} \alpha_i y_{n,m,t-i} + \sum_{j=0}^{b} \beta_j GDP_{n,t-j} + \sum_{k=0}^{c} \zeta_k PN_{n,t-k} + \sum_{l=0}^{d} \delta_l PO_{n,t-l} + \nu_t \quad (6.4)$$

$y_{n,m,t}$ represents the dependent variable in the model, which is the natural logarithm of the consumed amount of natural gas in region n, on consumption level m, in year t. $GDP_{n,t}$, $PN_{n,t}$ and $PO_{n,t}$ represents the independent variables in the model. These are the natural logarithms of GDP, the average natural gas price and the average oil price, respectively, all in region n, in year t. α_0 represents the intercept. ν_t represents the error-term.

Subscript n represents the geographical region of research, and it takes on the values 1, 2 and 3. These numbers represent OECD Europe, OECD Americas and OECD Asia & Oceania, respectively. Subscript m represents the level of natural gas consumption, and it takes on the descriptions *in*, *ov* and *si*. These descriptions stand for total industrial consumption, overall consumption and sub-industrial consumption, respectively. i, j, k and l are counting variables that separate the different terms in the model hailing from the various variables included. a, b, c and d represent the number of lag terms for each of the variables, respectively.

Subscript t is a vector $[1, ..., T]$ where T is the size of the sample utilized adjusted for the maximum amount of lags allowed in the model. All of the variables are vectors of size $(1 \times T)$. The same applies for ν_t. α_0 is a scalar of size (1×1). The value of T is the same in all cases analysed, except for Asia & Oceania - Industry where there are fewer data points available as discussed in Section 5.

The EC-parameterization of equation (6.4) is given in equation (6.5).

$$\begin{aligned} \Delta y_{n,m,t} = \alpha_0 + \sum_{i=1}^{a} \alpha_i^* \Delta y_{n,m,t-i} + \sum_{j=0}^{b} \beta_j^* \Delta GDP_{n,t-j} + \sum_{k=0}^{c} \zeta_k^* \Delta PN_{n,t-k} + \sum_{l=0}^{d} \delta_l^* \Delta PO_{n,t-l} \\ -\lambda \left(y_{t-1} - \Theta_{GDP} GDP_{n,t} - \Theta_{PN} PN_{n,t} - \Theta_{PO} PO_{n,t} \right) + \nu_t \end{aligned} \quad (6.5)$$

Here α_j^*, β_j^*, ζ_j^* and δ_j^* represent the various short-run coefficients for the consumption variable of choice, GDP, average natural gas price and the average oil price, respectively. The number of short-run coefficients for each variable depends on the number of terms the specific variable is represented by in the specific model. Θ_{GDP} Θ_{PN} and Θ_{PO} represent the long-run coefficients for the GDP, the average natural gas price and the average oil price, respectively.

The relevant exact equations are shown in Appendix A.

6.5 Diagnostics - Testing of the Models

In order to ensure that the Gauss-Markov Assumptions are upheld, and that the estimators used in the regression models are indeed the Best Linear Unbiased Estimators, a set of diagnostic tests has to be performed. The tests that will be applied in this thesis test for serial correlation, heteroskedasticity, model misspecification, normal distribution of the residuals and stability. Brief presentations of these tests are given below. As long as the chosen tests yield satisfying results for the respective regression models, the results from said models are deemed as valid.

Test for Serial Correlation

Serial correlation, also known as autocorrelation, is the event in which there is a proven relation between a variable and its lags. If the Gauss-Markov Assumptions are to be upheld, there can be no autocorrelation for the residuals in the specific model (Brooks, 2014). In order to test for serial correlation the Breusch-Godfrey Lagrange Multiplier test for autocorrelation (Godfrey, 1978) will be utilised. It tests for autocorrelation in the residuals, where the null hypothesis is that there is no serial correlation

Tests for Heteroskedasticity

Homoskedasticity is the term used when the variance of the residuals are zero. If this is not the case, the residuals are said to be heteroskedastic. If the residuals are heteroskedastic one of the Gauss-Markov Assumptions is not upheld (Brooks, 2014). In this thesis, the test for heteroskedasticity suggested by White (1980) will be applied. In this test, the null hypothesis is that the residuals are homoskedastic.

Test for Model Misspecification

Model misspecification tests revolves around identifying if the functional form of the model is correct, or if it is misspecified (Brooks, 2014). Ramsey (1969) formulates four tests for model misspecifications in linear least squares regression analysis. The regression specification-error test (RESET) is one of them, and will be applied in this thesis. It tests the effect omitted higher power variables have on the distribution of the residuals of the model. In essence the RESET tests if the suggested model should be linear or not (Brooks, 2014). The null hypothesis of the test is that there are no omitted variables.

Tests for Normal Distribution of the Residuals

For the residuals of a regression model to be normally distributed is in itself not one of the Gauss-Markov Assumptions. However it can have an impact on the validity of the calculations of the p-values that are used for testing the significance of the estimated coefficients. This applies especially in the cases where the size of the sample used is small, which is the case in this thesis. When the sample size is larger, a result of the Central Limit Theorem is that the residuals are approximately normally distributed (Brooks, 2014). As a result of this, the normality of the residuals will be tested in this thesis through testing the skewness and kurtosis of the distribution of the residuals.

Skewness is a term used to describe the shape of the distribution with respect to whether or not it is symmetrical about its mean. If a distribution has zero skewness, it is said to be normally distributed. Kurtosis is a term used to measure how ”fat” the tails of the distribution are. If a distribution has a kurtosis coefficient of three, it is said to be normally distributed (Brooks, 2014).

In this thesis, the test suggested by D'Agostino et al. (1990), which tests for skewness and kurtosis, jointly, will be applied. The null hypothesis of the test is normality.

In addition to these tests, visual inspection can be useful. Two types of plots that can be used for this purpose are the quantile-quantile plot and the normal-probability plot. In both cases if the plot forms a straight line with few outliers it indicates that the plotted sample is normally distributed. Both these plots will be applied in this thesis.

Tests of Stability

Testing for stability entails testing if the estimated coefficients are stable over time. This can be done through visual inspection by plotting the cumulative sum of the recursive residuals. A CUSUM-chart is a chart containing a plot of the cumulative sum of the recursive residuals as well as a set of significance lines, which represents the upper and lower confidence bounds of a chosen percentage confidence interval (Brooks, 2014). If the cumulative sum of the recursive residuals-plot breaches any of these lines, it is a sign of parameter-instability

In addition to this, the CUSUM-test introduced by Brown et al. (1975) can be applied. The null-hypothesis of this test is that the recursive residuals are independent and identically distributed (iid) as well as normally distributed with a mean of zero and constant variance. These are signs of the estimated coefficients being stable over time.

Both the CUSUM-test and the visual inspection will be conducted in this thesis.

6.6 Analysis of Results

In addition to the diagnostics, an interpretation and evaluation of the estimated coefficients will be conducted. In this lies that a discussion as to whether or not the estimates are logical with respect to magnitude and sign will be performed. This discussion will also touch upon whether or not they are in line with economic theory.

Furthermore, a discussion about the significance of the estimates with respect to their sensitivity to the lag-specification will be conducted. This is based on the argument made by Cuddington & Dagher (2015) which suggests that one should include at least two terms with different time subscripts for each variable in the model. This is suggested in order to avoid imposing a priori constraints that could possibly lead to implausible short- and long-run elasticities with respect to their magnitudes. As a result of this, all models suggested in this thesis that lacks at least two terms with different time subscripts for each variable will be tested against their counterparts that fulfil this requirement. This is in order to see if there are any significant differences between the different lag specifications.

6.7 Summary of Econometric Framework

A summary of the econometric method applied in this thesis is as follows:

1. Test for stationarity
2. Decide on lag specification
3. Estimate ARDL-coefficients
4. Test for Cointegration
5. Estimate short- and long-run coefficients if possible
6. Perform diagnostics tests
7. Conduct analysis of results

7 Results

In this section the results, using StataMP 15.1, from the empirical analysis as presented in Section 6 are presented. For the ARDL- and Error Correction-methods these are grouped based on the research questions from Section 1.

7.1 Stationarity and Lag Specification

As mentioned in Section 6, the variables used in this thesis must fulfil the requirement raised in relation to order of integration when applying the ARDL-framework. This entails all variables having orders of integration lower than 2, in order for the results hailing from the ARDL-models to be valid. To ensure that this requirement is met, the ADF test will be conducted such as presented in Subsection 6.2. The ADF test is conducted both in the event that a trend is present for the specific variable, as well as in the event where this is not the case. This is done in order to uncover possible significant differences this specification may have on the order of integration. This, alongside a visual inspection of the line plots of the chosen variables, serves as decision support with respect to the inclusion of the various variables in this thesis.

In order to decide on a lag structure, the SBIC presented in Subsection 6.3 is used. When applied to the various variables, the suggested number of lags included is one for all variables. Due to this, a lag of one is included in all the ADF tests conducted in this thesis.

The results of the ADF tests are presented in Tables 7 - 10. In these tables, the various test statistics and their corresponding p-values are presented alongside an interpretation. In Table 11 the Fuller (1996) Critical values used in this thesis are presented. These are determined based on the number of observations available of the variables used in this thesis, which was presented in Section 5.

Tables 7 - 10 show that all variables utilised in this thesis have an order of integration lower than 2. This is the case, irrespective of whether or not the ADF test is conducted including the presence of a trend. Furthermore, a visual inspection of the line plots of the level and first difference of the various variables, presented in Figures 4 - 17 in Appendix B.1, also indicate that the various variables all have an order of integration lower than 2.

While the presence of a trend in the ADF tests in some cases yield different interpretations than what is the case when a trend is not included, the specific order of integration is not important as long as the aforementioned requirement is met. Thus, the ARDL-framework can be utilised further.

	Level		First Difference		
ln(Cons. of NG)	**ADF**	**ADF - Trend**	**ADF**	**ADF - Trend**	**Interpretation**
Europe - Industry	-5.071 (0.000)	-4.418 (0.002)	-6.478 (0.000)	-6.701 (0.000)	I(0)**
Americas - Industry	-2.524 (0.110)	-2.648 (0.259)	-7.051 (0.000)	-7.092 (0.000)	I(1)**
Asia & Oceania - Industry	-2.970 (0.038)	0.523 (0.997)	-4.234 (0.001)	-5.099 (0.000)	I(0)*/I(1)**
Europe - Overall	-5.043 (0.000)	-3.937 (0.011)	-5.656 (0.000)	5.963 (0.000)	I(0)**/I(0)*
Europe - Sub-industry	-5.092 (0.000)	-5.086 (0.000)	-8.360 (0.000)	-8.294 (0.000)	I(0)**

Table 7: ADF Tests for Dependent Variables in different cases. The test-statistic is outside the parenthesis, and the corresponding MacKinnon p-value inside the parenthesis. (One asterisk indicates significance at 5 % level and two at 1 % level)

	Level		First Difference		
Ind. var. - Europe	**ADF**	**ADF - Trend**	**ADF**	**ADF - Trend**	**Interpretation**
ln(GDP)	-6.933 (0.000)	-2.826 (0.188)	-3.266 (0.017)	-4.739 (0.001)	I(0)**/I(1)**
ln(NG Price)	-2.758 (0.065)	-2.455 (0.351)	-5.790 (0.000)	-5.990 (0.000)	I(1)**
ln(Oil Price)	-3.054 (0.030)	-2.829 (0.186)	-5.939 (0.000)	-6.184 (0.000)	I(0)*/I(1)**

Table 8: ADF Tests for Independent Variables - OECD Europe. The test-statistic is outside the parenthesis, and the corresponding MacKinnon p-value inside the parenthesis. (One asterisk indicates significance at 5 % level and two at 1 % level)

	Level		First Difference		
Ind. var. - Americas	**ADF**	**ADF - Trend**	**ADF**	**ADF - Trend**	**Interpretation**
ln(GDP)	-8.975 (0.000)	-2.438 (0.360)	-2.820 (0.055)	-4.885 (0.000)	I(0)**/I(1)**
ln(NG Price)	-2.865 (0.050)	-1.920 (0.644)	-6.641 (0.000)	-7.301 (0.000)	I(0)*/I(1)**
ln(Oil Price)	-3.054 (0.030)	-2.829 (0.186)	-5.939 (0.000)	-6.184 (0.000)	I(0)*/I(1)**

Table 9: ADF Tests for Independent Variables - OECD Americas. The test-statistic is outside the parenthesis, and the corresponding MacKinnon p-value inside the parenthesis. (One asterisk indicates significance at 5 % level and two at 1 % level)

	Level		First Difference		
Ind. var. - Asia & Oceania	**ADF**	**ADF - Trend**	**ADF**	**ADF - Trend**	**Interpretation**
ln(GDP)	-9.577 (0.000)	-4.063 (0.007)	-2.852 (0.051)	-4.530 (0.001)	I(0)**
ln(NG Price)	-1.389 (0.588)	-1.621 (0.784)	-5.350 (0.000)	-5.275 (0.000)	I(1)**
ln(Oil Price)	-1.634 (0.466)	-1.865 (0.673)	-5.822 (0.000)	-5.767 (0.000)	I(1)**

Table 10: ADF Tests for Independent Variables - OECD Asia & Oceania. The test-statistic is outside the parenthesis, and the corresponding MacKinnon p-value inside the parenthesis. (One asterisk indicates significance at 5 % level and two at 1 % level)

Critical values - Level	**10 %**	**5 %**	**1 %**
Europe	-2.605	-2.941	-3.607
Europe - Trend	-3.190	-3.516	-4.187
Americas	-2.605	-2.941	-3.607
Americas - Trend	-3.190	-3.516	-4.187
Asia & Oceania	-2.612	-2.958	-3.648
Asia & Oceania - Trend	-3.204	-3.540	-4.242
Critical values - First Difference	**10 %**	**5 %**	**1 %**
Europe	-2.606	-2.944	-3.614
Europe - Trend	-3.192	-3.520	-4.196
Americas	-2.606	-2.944	-3.614
Americas - Trend	-3.192	-3.520	-4.196
Asia & Oceania	-2.613	-2.961	-3.655
Asia & Oceania - Trend	-3.206	-3.544	-4.251

Table 11: Fuller (1996) Critical Values for the Dickey & Fuller (1979) Unit Root Test (ADF Test)

While the SBIC applied on the separate variables yield the same result, this does not mean that an ARDL(1, 1, 1, 1)-model is preferable over other possible regression models. A different lag structure may in fact score better in terms of how the SBIC penalizes the use of 'resources', such as was mentioned in Subsection 6.3. As a result of this, multiple specifications have been tested. In the process of generating these specifications, various constraints have been put on the models. The aim of this approach have been to test what effect there is on the preferred lag specification when constraints are put on how many lags are allowed for the various variables.

A majority of the cases analysed lean towards lag specifications including zero or one lag of their respective variables. As a result of this, the regression models in this thesis are specified with a constraint which requires the maximum amount of lags per variable to be at most one. The lag structures that score the best with respect to the SBIC in the individual cases are selected and used further. Furthermore, what this constraint also entails is that the amount of observations from the various time series used are the same across the various cases analysed, except for the case revolving Asia & Oceania. The importance of this was discussed briefly in Section 5.

7.2 ARDL-results

In Tables 12 and 13, the results hailing from the various regression models are presented. The results are grouped in the two tables based on how the research questions in this thesis were formulated. What is worth noting is that in Table 12, which presents the models representing Industry level natural gas consumption in the three chosen regions, all models have taken different lag structures. This stands in contrast to the models presented in Table 13, which represents three levels of natural gas consumption, all in Europe. For these cases, all three models have taken the same lag structure.

Due to the double log specifications, the estimated coefficients are interpreted as percentage increases or decreases. An example from Table 13 is the estimated coefficient for GDP for **Europe - Industry**, which indicates that a 1 % increase in GDP is associated with a 7.5 % increase in the industry level natural gas consumption, ceteris paribus. This interpretation holds for all of the estimated coefficients in the various regression models as long as they are statistically significant.

Model	**Europe - Industry**	**Americas - Industry**	**Asia & Oceania - Industry**
	ARDL(1, 0, 0, 1)	ARDL(1, 1, 0, 0)	ARDL(1, 0, 1, 0)
Ind. var. / Dep. var.	ln(Cons. of NG)	ln(Cons. of NG)	ln(Cons. of NG)
ln(Cons. of NG)$_{t-1}$	0.731** (0.046)	0.558** (0.101)	0.755** (0.076)
ln(GDP)$_t$	0.075** (0.013)	2.191** (0.458)	0.209** (0.070)
ln(GDP)$_{t-1}$	-	-2.117** (0.441)	-
ln(NG Price)$_t$	-0.043 (0.037)	-0.054* (0.020)	0.078 (0.055)
ln(NG Price)$_{t-1}$	-	-	-0.091** (0.030)
ln(Oil Price)$_t$	0.067* (0.026)	0.002 (0.017)	0.007 (0.030)
ln(Oil Price)$_{t-1}$	-0.076** (0.016)	-	-
Constant	2.295** (0.477)	4.457** (1.390)	-0.680 (0.417)

Table 12: ARDL-results - Part 1. The estimated coefficient is outside the parenthesis, and the corresponding standard error inside the parenthesis. (One asterisk indicates significance at 5 % level and two at 1 % level)

Model	Europe - Industry	Europe - Overall	Europe - Sub-industry
	ARDL(1, 0, 0, 1)	ARDL(1, 0, 0, 1)	ARDL(1, 0, 0, 1)
Ind. var. / Dep. var.	ln(Cons. of NG)	ln(Cons. of NG)	ln(Cons. of NG)
ln(Cons. of NG)$_{t-1}$	0.731** (0.046)	0.828** (0.051)	0.483** (0.094)
ln(GDP)$_t$	0.075** (0.013)	0.093**(0.025)	0.044** (0.015)
ln(GDP)$_{t-1}$	-	-	-
ln(NG Price)$_t$	-0.043 (0.037)	0.050 (0.040)	-0.087 (0.056)
ln(NG Price)$_{t-1}$	-	-	-
ln(Oil Price)$_t$	0.067* (0.026)	-0.012 (0.029)	0.113** (0.040)
ln(Oil Price)$_{t-1}$	-0.076** (0.016)	-0.074** (0.017)	-0.068** (0.024)
Constant	2.295** (0.477)	0.612 (0.391)	5.058** (0.963)

Table 13: ARDL-results - Part 2. The estimated coefficient is outside the parenthesis, and the corresponding standard error inside the parenthesis. (One asterisk indicates significance at 5 % level and two at 1 % level)

7.3 Cointegration

In order to prove whether or not any long-run relationships exist the Pesaran et al. (2001) bounds test using the Kripfganz & Schneider (2018) finite-sample and asymptotic bounds is applied such as mentioned in Section 6.3. The F- and t-statistics of the various cases are presented in Table 14, alongside interpretations based on the two separate bounds applied. The specific bounds for the various cases are presented in Table 15.

For **Europe - Industry**, **Americas - Industry** and **Europe - Sub-industry**, both the F- and t-statistics are outside their respective bounds of at least a 5 % significance level, meaning that they pass the Pesaran et al. (2001) bounds test. In the cases of **Asia & Oceania - Industry** and **Europe - Overall**, the F-statistics are outside their respective bounds. However, this is not the case for the t-statistics. Thus, they do not pass the Pesaran et al. (2001) bounds test, and as a result of this, no EC-reparameterizations are performed for these two cases.

	F-Statistic	t-Statistic	Finite Sample	Asymptotic
Europe - Industry	20.907	-5.801	Yes**	Yes**
Americas - Industry	9.707	-4.385	Yes*	Yes**
Asia & Oceania - Industry	5.661	-3.228	No	No
Europe - Overall	19.204	-3.340	No	No
Europe - Sub-industry	9.515	-5.471	Yes**	Yes**

Table 14: Test-statistics for the Pesaran et al. (2001) Bounds test - Case 3, and Evaluation using Kripfganz & Schneider (2018) Critical Values (One asterisk indicates significance at 5 % level, and two asterisks at 1 % level.)

Finite Sample - F-Statistic	**10 %**	**5 %**	**1 %**
Europe - Industry	[2.874 , 3.986]	[3.478 , 4.714]	[4.875 , 6.373]
Americas - Industry	[2.874 , 3.986]	[3.478 , 4.714]	[4.875 , 6.373]
Asia & Oceania - Industry	[2.907 , 4.038]	[3.534 , 4.802]	[5.004 , 6.569]
Europe - Overall	[2.874 , 3.986]	[3.478 , 4.714]	[4.875 , 6.373]
Europe - Sub-industry	[2.874 , 3.986]	[3.478 , 4.714]	[4.875 , 6.373]
Finite Sample - t-Statistic	**10 %**	**5 %**	**1 %**
Europe - Industry	[-2.568 , -3.455]	[-2.901 , -3.827]	[-3.569 , -4.562]
Americas - Industry	[-2.568 , -3.455]	[-2.901 , -3.827]	[-3.569 , -4.562]
Asia & Oceania - Industry	[-2.570 , -3.458]	[-2.910 , -3.842]	[-3.598 , -4.606]
Europe - Overall	[-2.568 , -3.455]	[-2.901 , -3.827]	[-3.569 , -4.562]
Europe - Sub-industry	[-2.568 , -3.455]	[-2.901 , -3.827]	[-3.569 , -4.562]
	10 %	**5 %**	**1 %**
Asymptotic - F-Statistic	[2.729 , 3.747]	[3.226 , 4.321]	[4.296 , 5.535]
	10 %	**5 %**	**1 %**
Asymptotic - t-Statistic	[-2.569 , -3.426]	[-2.864 , -3.744]	[-3.434 , -4.342]

Table 15: Kripfganz & Schneider (2018) Critical Values for the Pesaran et al. (2001) Bounds Test - Case 3

7.4 EC-results

In Tables 16 and 17, the results hailing from the EC-reparameterizations of the various regression models are presented. The long-run coefficients are listed in level form. The short-run coefficients, on the other hand, are listed in first difference form. Similar to the ARDL-results, the coefficients represent a certain percentage increase or decrease in the level of natural gas consumption due to a percentage increase or decrease of the specific variable. Furthermore, the speed of adjustment coefficients are in all three cases negative and significant, which is a requirement for the models to converge towards equilibrium, such as was discussed in Section 6.3.

Model	Europe - Industry	Americas - Industry	Asia & Oceania - Industry
	ARDL(1, 0, 0, 1)	ARDL(1, 1, 0, 0)	ARDL(1, 0, 1, 0)
Ind. var. / Dep. var.	Δln(Cons. of NG)	Δln(Cons. of NG)	Δln(Cons. of NG)
Speed of Adjustment	-0.269** (0.046)	-0.442** (0.101)	-
Long-run			
ln(GDP)	0.277** (0.038)	0.167* (0.076)	-
ln(NG Price)	-0.158 (0.135)	-0.123* (0.047)	-
ln(Oil Price)	-0.034 (0.112)	0.005 (0.038)	-
Short-run			
$\Delta\ln(\text{GDP})_t$	-	2.117** (0.441)	-
$\Delta\ln(\text{NG Price})_t$	-	-	-
$\Delta\ln(\text{Oil Price})_t$	0.076** (0.016)	-	-
Constant	2.295** (0.477)	4.457** (1.390)	-

Table 16: EC-results - Part 1. The estimated coefficient is outside the parenthesis, and the corresponding standard error inside the parenthesis. (One asterisk indicates significance at 5 % level and two at 1 % level)

Model	Europe - Industry	Europe - Overall	Europe - Sub-industry
	ARDL(1, 0, 0, 1)	ARDL(1, 0, 0, 1)	ARDL(1, 0, 0, 1)
Ind. var. / Dep. var.	Δln(Cons. of NG)	Δln(Cons. of NG)	Δln(Cons. of NG)
Speed of Adjustment	-0.269** (0.046)	-	-0.517** (0.094)
Long-run			
ln(GDP)	0.277** (0.038)	-	0.085** (0.029)
ln(NG Price)	-0.158 (0.135)	-	-0.168 (0.112)
ln(Oil Price)	-0.034 (0.112)	-	0.087 (0.089)
Short-run			
$\Delta\ln(\text{GDP})_t$	-	-	-
$\Delta\ln(\text{NG Price})_t$	-	-	-
$\Delta\ln(\text{Oil Price})_t$	0.076** (0.016)	-	0.068** (0.024)
Constant	2.295** (0.477)	-	5.058** (0.963)

Table 17: EC-results - Part 2. The estimated coefficient is outside the parenthesis, and the corresponding standard error inside the parenthesis. (One asterisk indicates significance at 5 % level and two at 1 % level)

7.5 Diagnostics

In Table 18 the test statistics for the tests presented in Subsection 6.5 and applied on the various cases analysed in this thesis are presented. In none of the tests, the null hypothesis is rejected. In addition to these tests, quantile-quantile and normal-probability plots are presented in Appendix B.2 and B.3. Based on their appearances, all show signs that the residuals are normally distributed. As a result of this, and the test-statistics shown in Table 18, the estimators utilised in this thesis can be viewed as being the Best Linear Unbiased Estimators, such as discussed in Section 6.5.

In Table 19 the test statistics for the Brown et al. (1975) CUSUM test are presented. The critical values for the test are presented in Table 20. The test statistics in Table 19 indicate that the null hypothesis is not rejected in any of the cases analysed in this thesis. This is supported through a visual inspection of the corresponding CUSUM plots presented in Appendix B.4. In the CUSUM plot for **Americas - Industry**, shown in Figure 29, a minor breach of the boundaries is observed. The plot does however stabilize after this breach. Thus, the model can be viewed as stable over the period as a whole.

Since all the cases analysed in this thesis yield satisfying results with respect to the diagnostics tests presented, the results from these regression models are deemed as valid.

Statistic	**Europe - Industry**	**Americas - Industry**	**Asia & Oceania - Industry**	**Europe - Overall**	**Europe - Sub-industry**
Serial Correlation - (Godfrey, 1978)	2.046 (0.1526)	1.139 (0.2859)	0.224 (0.6362)	0.692 (0.4054)	0.685 (0.4079)
Heteroskedasticity - (White, 1980)	24.59 (0.2175)	19.22 (0.5078)	21.84 (0.3493)	24.53 (0.2198)	26.31 (0.1557)
Model Misspecification - (Ramsey, 1969)	0.18 (0.9095)	0.82 (0.4933)	1.51 (0.2314)	0.50 (0.6818)	1.37 (0.2665)
Normality - (D'Agostino et al., 1990)	4.05 (0.1318)	2.47 (0.2903)	0.83 (0.6608)	2.93 (0.2315)	2.09 (0.3526)

Table 18: Diagnostic Tests for the ARDL-models. The test statistic is shown outside the parenthesis, and the corresponding p-value inside the parenthesis.

Statistic	**Europe - Industry**	**Americas - Industry**	**Asia & Oceania - Industry**	**Europe - Overall**	**Europe - Sub-industry**
Stability (Brown et al., 1975)	0.6823	0.9543	0.5734	0.5512	0.7323

Table 19: Brown et al. (1975) CUSUM Test for the ARDL-models

Critical values	**10 %**	**5 %**	**1 %**
	0.850	0.9479	1.1430

Table 20: Critical Values for the Brown et al. (1975) CUSUM Test

8 Interpretation, Discussion, Limitations and Extensions

This section contains an analysis of the results presented in Section 7. In this analysis the results are interpreted and discussed both with respect to existing literature and to the research questions raised in Section 1. Furthermore a discussion on the scientific method utilised in this thesis is done. This touches upon the quality of the work done, as well as some limitations. Last, some possible extensions of the regression models presented in this thesis are proposed.

8.1 Analysis of Results

8.1.1 General Analysis of Results

When looking at the ARDL-results in Tables 12 and 13, what is evident is that a majority of the estimated coefficients in the various models presented are significant at 5 % level. This does, however, not apply in the cases of **Asia & Oceania - Industry** and **Europe - Overall**, where half of the estimated coefficients are not statistically significant. Apart from the fact that a coefficient being statistically insignificant implies that one cannot exclude the possibility that it may take the value of 0, the statistically insignificant coefficients will not be discussed further.

When comparing only the statistically significant coefficients from the various models it is evident that most share the same sign and a similar magnitude.

In the cases of the coefficients corresponding to independent variables with only one term, all signs correspond with economic theory. The coefficients related to GDP are positive, the ones related to own price are negative and the ones related to the cross price are positive.

In the cases of the coefficients corresponding to independent variables with two terms, all lag coefficients have the opposite sign of their level counterpart if they are both statistically significant. This applies to all the cases with pairs of coefficients except for **Europe - Overall**. In this specific case, only one of the two oil related coefficients is statistically significant. In addition, the sign of this coefficient is negative, which in itself is not in line with economic theory. However, since it is the lag coefficient that is statistically significant one should note that it shares the same sign as the other lagged oil related coefficients.

If the coefficients are interpreted as income, own price and cross price elasticities of demand, one could, based on their magnitudes, claim that all are relatively inelastic. This would mean that the level of consumption is not affected quickly by changes in the various independent variables. The only exception to this is the two coefficients related to GDP for **Americas - Industry**. What is worth noting, however, is that the overall impact of these two coefficients nearly cancel each other out. As a result of this, one could interpret the level of industry level natural gas consumption in the OECD Americas as being relatively inelastic to the GDP, based on these results.

The same interpretation can be applied when looking at the cases with pairs of coefficients with opposite signs mentioned above. These seem to cancel each other out, with the overall sign from the variable being in line with economic theory. The only exception to this is in the case of the coefficients related to the oil price for **Europe - Industry**. In this case, the magnitude of the negative coefficient is greater than its positive counterpart.

When comparing this to the long-run and short-run coefficients for this variable in the EC-results presented in Tables 16 and 17, one can see that only the short run coefficient is statistically significant. Furthermore, the sign is positive, which is more in line with what one would expect in a cross price elasticity of demand.

When looking at the remaining EC-coefficients, all of the long-run coefficients for GDP are statistically significant. If interpreted as income elasticities of demand their magnitudes imply that they are all rather inelastic. Among the other long-run coefficients, only the **Americas - Industry** natural gas price coefficient is statistically significant. Also this is fairly inelastic when interpreted as an own-price elasticity of natural gas demand. In addition to this, the signs of the various long-run coefficients are in line with economic theory.

All short-run coefficients are statistically significant and similar to their counterparts in the ARDL-results. The only short-run coefficient that stands out is the coefficient for GDP for **Americas - Industry**. If interpreted as an income elasticity of demand it is fairly elastic. This stands in contrast with the other coefficients, which are relatively inelastic. Furthermore, what this indicates is that the discussion made above about the corresponding ARDL-GDP coefficients possibly cancelling each other out, might not be the case after all.

Comparing the results presented in this thesis with the literature reviewed in Section 4 what seems to be the case is that the signs and magnitudes of the various coefficients presented in the literature for the most part are in line with the findings in this thesis. These are in turn in line with economic theory.

As mentioned in Section 4 most of the papers reviewed only look at own price and income elasticities. The few cases where a cross price elasticity of demand based on oil prices have been estimated, the results match the findings in this thesis with respect to sign and magnitude, in that all are positive and all are inelastic. Regarding own price elasticities most are inelastic and negative which also match the findings in this thesis.

Looking at income elasticities all are positive. This is similar to the findings in this thesis. When it comes to magnitude approximately half of all the coefficients related to income/GDP are elastic. The other half is inelastic. Thus the findings in this paper regarding the income elasticity for **Americas - Industry** is not an unusual finding. On the other hand, what is more unusual is that the long run income elasticity in this case is inelastic, while the short run counterpart is elastic. Normally one would expect it to be the other way around, which for instance was the finding in Erdogdu (2010). The differences in short and long run elasticities are however covered to a varying degree in the other papers reviewed for this thesis. This goes both with respect to it being a focus, and with respect to the presence of significant coefficients. As a result of this it can be an interesting aspect to investigate further in future work within the field.

8.1.2 Importance of Lag Structure

When deciding on lag specifications, the results from the regression models with the maximum lag constraint previously described have been tested against the results from the same models where the lag structures have been forced to include exactly one lag per variable. The reason for this was to investigate if there were any significant differences in the two alternative models for each case. The rationale behind performing this test is a train of argumentation brought forward in Cuddington & Dagher (2015), which was discussed in Section 6.6. They suggest to include at least two terms with different time subscripts for each independent variable in the regression model. Thus, full empirical analyses of all five cases with ARDL(1, 1, 1, 1) specifications have been performed. The results are presented in Appendix C alongside the results of the corresponding Cointegration test and Diagnostic tests. Quantile-quantile, normal-probability and CUSUM plots are presented in Appendix B, next to their counterparts hailing from the models based on the maximum lag constraint.

When comparing the results of the two alternative regression models for the various cases, no significant differences with respect to the magnitudes and signs of the statistically significant coefficients are found. However, fewer coefficients are significant when using the ARDL(1, 1, 1, 1) lag specification. As a result of this, the lag specifications hailing from the alternative with the maximum lag constraint are kept as the preferred specifications, and the above analysis stands firm. This is due to there being more significant coefficients in this alternative, and also because these models are more parsimonious than the models using specifications based on the suggestion from Cuddington & Dagher (2015).

8.1.3 Summary of Analysis with respect to the Research Questions

Relating these findings back to the research questions raised in Section 1 it is evident that there for the most part are similarities in the relationships between the level of natural gas consumption and the various independent variables, applied in this thesis, when it comes to sign and magnitude. This is the case both for the level of industrial consumption across the three different regions, and for the various levels of consumption within OECD Europe. There are however a few exceptions.

With respect to the sign of the various coefficients, all the statistically significant coefficients have signs that are in line with economic theory when you only look at the EC-results. This is not the case when you look at the ARDL-results. What is worth mentioning is however, that all the discrepancies related to this can be associated with the cases where there are multiple terms associated with specific independent variables. This applies to all the cases analysed, meaning it is not a unique feature in any of the five cases.

When it comes to the magnitude of the various coefficients, all the statistically significant coefficients are relatively inelastic when interpreted as elasticities of demand. The only exception is the coefficients related to GDP for **Americas - Industry**, in which the ARDL-results are shown to be elastic. The EC-reparameterization, on the other hand, shows that this is the case only in the short run. In the long run it is inelastic, like all the other coefficients estimated in this thesis.

8.2 Discussion of Scientific Method

8.2.1 Quality of Quantitative Work

In order to assess the quality of a quantitative study, Heale & Twycross (2015), highlights two specific measures; validity and reliability.

According to Heale & Twycross (2015), when discussing the validity of a quantitative study, the focus lies on whether or not the aspects studied are correct. When applied to this thesis, such a discussion would revolve around whether or not the choices of variables and geographical regions make sense. As mentioned in Section 5, the rationale behind the choice of variables is the results of the literature review conducted. Furthermore, the choice of the geographical areas was based on the fact that the markets for natural gas are less integrated compared to the markets of resources with similar applications, such as for instance oil. Thus, the regions of choice represent some of the major natural gas markets as a whole. Based on this, one could argue that the aspects studied in this thesis are valid.

One could argue that the regression models are very general, compared to similar studies. This is, however, the result of choosing regions instead of smaller geographical areas. Finding independent variables that can be applied on a regional basis, might be a problem. Also, it can be difficult to find variables that have a higher sample frequency when the geographical region is larger. Thus, given the boundaries chosen, the approach is okay.

For the work to be reliable, one have to assess whether or not the results presented can be reproduced. All the data used in this thesis have been acquired through data bases that are available through a license acquired through the Norwegian University of Science and

Technology (NTNU). Thus it is highly likely that other researchers will be able to access the same data. Apart from a log transformation, nothing has been done to the data. Furthermore, the econometric framework applied is well known. The same goes for the software utilised. Based on this, it is reasonable to believe that the work presented in this thesis can be reproduced.

8.2.2 Limitations of the Analysis

While the approach chosen in this thesis yields statistically significant coefficients it is important to note that it is not possible to deduct which variables that have the highest impact from these coefficients alone. In order to rank the variables based on their impact on the dependent variable, other approaches are more appropriate.

This aspect is also relevant when attempting to compare resulting coefficients across the various cases analysed in this thesis. Based on the approach chosen in this thesis, it is not possible to assess the differences in what impact various independent variables within a specific model have on the dependent variable. Furthermore, due to the models having somewhat differing lag specifications, in addition to the sample size being slightly smaller in one of the cases, one should not draw full conclusions based on the results presented in this thesis. What is possible, however, is to discuss the magnitudes and signs of the estimated coefficients and see if there are any similarities or differences in their individual interpretations, which is the approach chosen in this thesis.

8.3 Generalizability and Possible Further Work

In addition to addressing the validity and reliability of a study alongside the limitations of the analysis, it is important to discuss the generalizability of the results. Mills et al. (2010) highlights this. The aspect of generalizability revolves around whether or not the results can be generalized, based on the samples utilised. While the results presented in this thesis make practical sense, one should be careful claiming that they are generalizable given that they are based on aggregated variables for aggregated regions. A solution to this could be to analyse the individual countries of the various regions in order to see if their results are similar to their aggregate counterparts. This was beyond the scope of this thesis, but is interesting, nonetheless. A reason for this is that a risk associated with using aggregated data is that you end up with some type of middle value, that might not represent any of the underlying areas. This could be the case if the region of research comprises of smaller areas that when analysed empirically have diametrically different coefficient values. An extension of the work done in this thesis could thus be to investigate this aspect in order to assess the generalizability of empirical analyses of regional natural gas consumption.

Another aspect that could be investigated is performing a similar analysis, in the same time period, only with a larger sample size. This can be achieved through increasing the sample frequency, for instance by using quarterly or monthly data. Doing so can contribute to reduce the risks associated with smaller sample sizes. Since the regression models in this thesis targets aggregated regions, using more high frequency data can be challenging, as it depends highly on what sort of data is made available. If this kind of data is available, however, it could be interesting to compare the two kinds of models in order to see if there are any significant differences.

Furthermore, in this thesis, none of the models have been adjusted based on specific extraordinary events that could have had an impact on the values of the variables. Examples of such events could be the 9/11 terrorist attacks, the 2008 financial crisis or the 1973 oil crisis. This could have been done through implementing dummy variables. The rationale behind keeping the models as they were was to make them as comparable as possible. A further extension of the work done in this thesis could be to take such events into account, and evaluate to what extent the various models are affected compared to what was analysed in this thesis.

8.4 An Extension including Carbon Prices

In continuation of the discussion of possible further work, it would be interesting to investigate the relationship between the level of natural gas consumption and carbon prices. As highlighted in Section 1 there is a large global consensus that there is a need for a decrease in the levels of both consumption of non renewables and climate gas emissions. Over the past couple of decades several measures have been discussed and implemented with respect to this.

Among the measures implemented is the EU Emission Trading Scheme (ETS) which is one of the major tools used for targeting climate gas emissions in Europe. The EU ETS is a cap and trade system where companies can trade emission quotas (European Commision, 2015). Within this system a dynamic price on carbon emission has been established, where the purpose of this is fining emissions that exceed the yearly quota of a given company. This system has been up and running since 2005, meaning that annual carbon price data dates back to then.

Not only does this scheme facilitate the direct reduction of climate gas emission, but it also stimulates investment towards the implementation of "green technologies" that are more efficient with respect to emissions (European Commision, 2015). An example of such a technology is Carbon Capture Utilisation and Storage (CCUS). According to IEA (2019b) this is a technology that hopefully can help decarbonize the industry sector, which comprises of several energy intensive sub industries, such as mentioned in Section 2. One of them is the Chemicals and Petrochemicals industry which is the sub-industry analysed in this thesis.

Based on this, the inclusion of carbon prices as an independent variable would be an interesting extension to the regression models presented in this thesis. An attempt was made at including the EU ETS prices as a proxy for carbon pricing in the various cases associated with OECD Europe in this thesis. Unfortunately, the inclusion of annual EU ETS prices did not yield satisfactory results. The main reason for this was the small amount of data points available. Because of the small amount of data points, the remaining variables had to be adjusted so that they were all of equal length. What this did was change the properties of the remaining variables with respect to stationarity. As a result of this, the requirements for order of integration associated with the use of the ARDL-framework were no longer upheld. Thus, none of the resulting coefficients from these models could be deemed valid.

A solution to this is simply to wait until more data points are made available. This will unfortunately take several years. It would, however, make an interesting addition to natural gas models as they are known today.

9 Conclusion

While there is political consensus throughout the world that the level of consumption of non renewables needs to decrease over time, this is in some cases difficult to achieve. In the industry sector, for instance, natural gas plays a role that is currently hard to replace. Thus, a demand for natural gas will exist for many years to come. As a result of this it is of interest to gain a good understanding of what determines the demand of natural gas, and how the relationships between these determinants and the level of natural gas consumption may differ throughout the world.

In this thesis, a set of cases are empirically analysed using an ARDL-framework and OLS estimators. The purpose of analysing these cases is to investigate if there are any differences in the relationships between specific levels of natural gas consumption throughout the world and some of the most commonly used determinants of natural gas demand; GDP, natural gas prices and oil prices.

First, the industry level of natural gas consumption is analysed for three specific geographical regions, representing the three main markets for natural gas; OECD Europe, OECD Americas and OECD Asia & Oceania. As the markets for natural gas are less integrated compared to the markets of comparable commodities, such as oil, the aim is to investigate if there are any significant differences in the aforementioned relationships **across** these regions.

Second, the overall and sub-industrial level of natural gas consumption of OECD Europe are analysed and compared to its industrial counterpart, using the same set of determinants of demand. The purpose of this is to investigate if there are any differences in the aforementioned relationships **within** the same region. The sub-industry of choice is the chemical and petrochemical industry. This choice was made due to it being one of the sub-industries where it is deemed difficult to switch away from using natural gas.

In addition to the ARDL-models, the five cases are tested for cointegration. If the presence of cointegration is proven in a specific case, an EC-reparameterization is conducted. The purpose of this is to split the aforementioned relationships into long-run and short-run parts.

The resulting coefficients of the various regression models show similarities in the signs and magnitudes. When using these as measures to describe the aforementioned relationships it is apparent that there are clear similarities across the various cases analysed. For the cases focusing on industry level of natural gas consumption there are few differences across the various regions. The same can be said regarding the various levels of natural gas consumption analysed within OECD Europe. Across the five cases analysed, there are a couple of coefficients with magnitudes that differ from the rest. However, comparing this to the findings in the existing literature reviewed in relation to this thesis, these deviations are not unique.

The work done in this thesis adds to the field of research in that it lifts the level of focus in several ways. While the approach of empirically analysing different levels of natural gas consumption is not new, it has seldom been done for larger regions. Furthermore, there are few papers focusing on making comparisons across different geographical areas. What the results of this thesis show is that even though the level of focus in this thesis is aggregated regions, there are similarities with respect to signs and magnitudes when making comparisons to the results in papers with more disaggregated levels of focus.

If the purpose of empirical analysis of natural gas demand is to better understand the nature of the relationships between the level of consumption and the various determinants of demand, it is of interest to make comparisons across different scopes to see if the results can be generalised. While insight into smaller cases is also useful and important, being able to generalise them to apply for larger areas opens up for the creation of measures and policies that can have a wider field of impact.

While the regression models presented in this thesis are very general they still provide insight in that they contribute to generalise the findings of existing literature which focus on smaller areas. An extension of the work done in this thesis could be to attempt to make the various models more nuanced. This can be achieved through adjusting them based on specific extraordinary events in history. This can also be done through the inclusion of additional variables, for instance carbon prices.

A Appendix: Equations

Equations (A.1) - (A.10) represent the general ARDL-specifications and their corresponding EC-formulations for the five cases that are analysed in this thesis.

OECD Europe - Industry

$$y_{1,in,t} = \alpha_0 + \sum_{i=1}^{a} \alpha_i y_{1,in,t-i} + \sum_{j=0}^{b} \beta_j GDP_{1,t-j} + \sum_{k=0}^{c} \zeta_k PN_{1,t-k} + \sum_{l=0}^{d} \delta_l PO_{1,t-l} + \nu_t \quad (A.1)$$

$$\Delta y_{1,in,t} = \alpha_0 + \sum_{i=1}^{a} \alpha_i^* \Delta y_{1,in,t-i} + \sum_{j=0}^{b} \beta_j^* \Delta GDP_{1,t-j} + \sum_{k=0}^{c} \zeta_k^* \Delta PN_{1,t-k} + \sum_{l=0}^{d} \delta_l^* \Delta PO_{1,t-l}$$
$$-\lambda \left(y_{1,in,t-1} - \Theta_{GDP} GDP_{1,t} - \Theta_{PN} PN_{1,t} - \Theta_{PO} PO_{1,t}\right) + \nu_t \quad (A.2)$$

OECD Americas - Industry

$$y_{2,in,t} = \alpha_0 + \sum_{i=1}^{a} \alpha_i y_{2,in,t-i} + \sum_{j=0}^{b} \beta_j GDP_{2,t-j} + \sum_{k=0}^{c} \zeta_k PN_{2,t-k} + \sum_{l=0}^{d} \delta_l PO_{2,t-l} + \nu_t \quad (A.3)$$

$$\Delta y_{2,in,t} = \alpha_0 + \sum_{i=1}^{a} \alpha_i^* \Delta y_{2,in,t-i} + \sum_{j=0}^{b} \beta_j^* \Delta GDP_{2,t-j} + \sum_{k=0}^{c} \zeta_k^* \Delta PN_{2,t-k} + \sum_{l=0}^{d} \delta_l^* \Delta PO_{2,t-l}$$
$$-\lambda \left(y_{2,in,t-1} - \Theta_{GDP} GDP_{2,t} - \Theta_{PN} PN_{2,t} - \Theta_{PO} PO_{2,t}\right) + \nu_t \quad (A.4)$$

OECD Asia & Oceania - Industry

$$y_{3,in,t} = \alpha_0 + \sum_{i=1}^{a} \alpha_i y_{3,in,t-i} + \sum_{j=0}^{b} \beta_j GDP_{3,t-j} + \sum_{k=0}^{c} \zeta_k PN_{3,t-k} + \sum_{l=0}^{d} \delta_l PO_{3,t-l} + \nu_t \quad (A.5)$$

$$\Delta y_{3,in,t} = \alpha_0 + \sum_{i=1}^{a} \alpha_i^* \Delta y_{3,in,t-i} + \sum_{j=0}^{b} \beta_j^* \Delta GDP_{3,t-j} + \sum_{k=0}^{c} \zeta_k^* \Delta PN_{3,t-k} + \sum_{l=0}^{d} \delta_l^* \Delta PO_{3,t-l} - \lambda \left(y_{3,in,t-1} - \Theta_{GDP} GDP_{3,t} - \Theta_{PN} PN_{3,t} - \Theta_{PO} PO_{3,t} \right) + \nu_t \quad \text{(A.6)}$$

OECD Europe - Overall

$$y_{1,ov,t} = \alpha_0 + \sum_{i=1}^{a} \alpha_i y_{1,ov,t-i} + \sum_{j=0}^{b} \beta_j GDP_{1,t-j} + \sum_{k=0}^{c} \zeta_k PN_{1,t-k} + \sum_{l=0}^{d} \delta_l PO_{1,t-l} + \nu_t \quad \text{(A.7)}$$

$$\Delta y_{1,ov,t} = \alpha_0 + \sum_{i=1}^{a} \alpha_i^* \Delta y_{1,ov,t-i} + \sum_{j=0}^{b} \beta_j^* \Delta GDP_{1,t-j} + \sum_{k=0}^{c} \zeta_k^* \Delta PN_{1,t-k} + \sum_{l=0}^{d} \delta_l^* \Delta PO_{1,t-l} - \lambda \left(y_{1,ov,t-1} - \Theta_{GDP} GDP_{1,t} - \Theta_{PN} PN_{1,t} - \Theta_{PO} PO_{1,t} \right) + \nu_t \quad \text{(A.8)}$$

OECD Europe - Sub-industry

$$y_{1,si,t} = \alpha_0 + \sum_{i=1}^{a} \alpha_i y_{1,si,t-i} + \sum_{j=0}^{b} \beta_j GDP_{1,t-j} + \sum_{k=0}^{c} \zeta_k PN_{1,t-k} + \sum_{l=0}^{d} \delta_l PO_{1,t-l} + \nu_t \quad \text{(A.9)}$$

$$\Delta y_{1,si,t} = \alpha_0 + \sum_{i=1}^{a} \alpha_i^* \Delta y_{1,si,t-i} + \sum_{j=0}^{b} \beta_j^* \Delta GDP_{1,t-j} + \sum_{k=0}^{c} \zeta_k^* \Delta PN_{1,t-k} + \sum_{l=0}^{d} \delta_l^* \Delta PO_{1,t-l} - \lambda \left(y_{1,si,t-1} - \Theta_{GDP} GDP_{1,t} - \Theta_{PN} PN_{1,t} - \Theta_{PO} PO_{1,t} \right) + \nu_t \quad \text{(A.10)}$$

B Appendix: Figures

B.1 Data

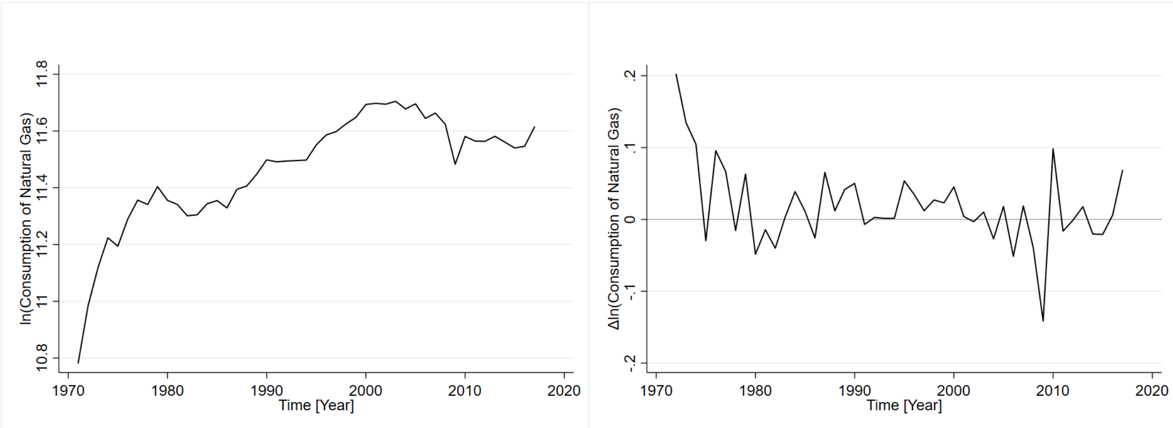


Figure 4: Level and First Difference Plots of Natural Gas Consumption - Europe - Industry. Source: OECD iLibrary

Figure 5: Level and First Difference Plots of Natural Gas Consumption - Americas - Industry. Source: OECD iLibrary

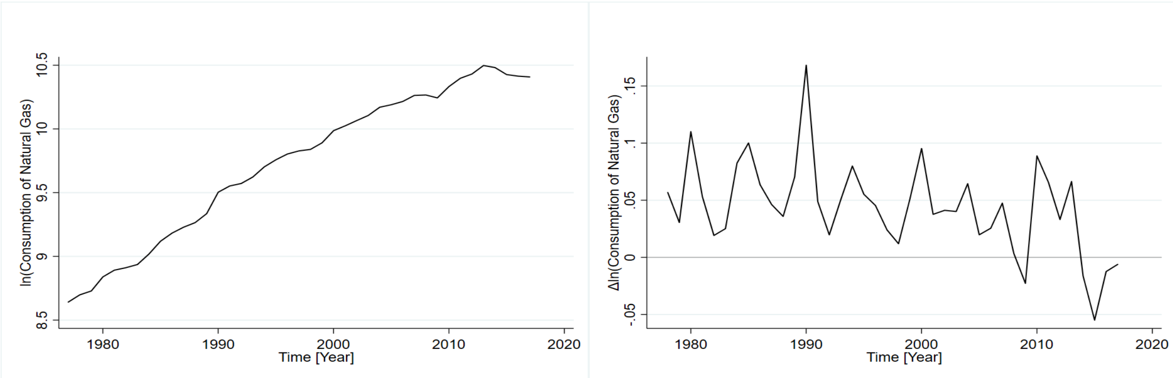


Figure 6: Level and First Difference Plots of Natural Gas Consumption - Asia & Oceania - Industry. Source: OECD iLibrary

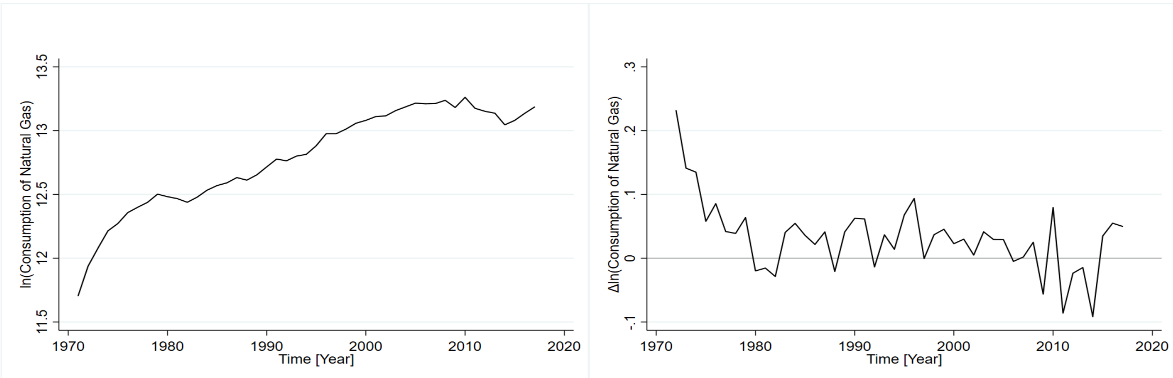


Figure 7: Level and First Difference Plots of Natural Gas Consumption - Europe - Overall. Source: OECD iLibrary

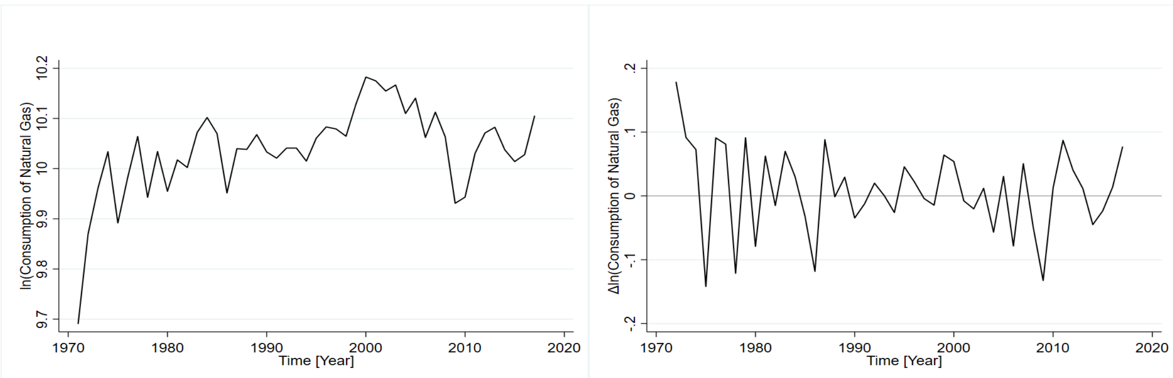


Figure 8: Level and First Difference Plots of Natural Gas Consumption - Europe - Sub-industry. Source: OECD iLibrary

B.1.1 OECD Europe

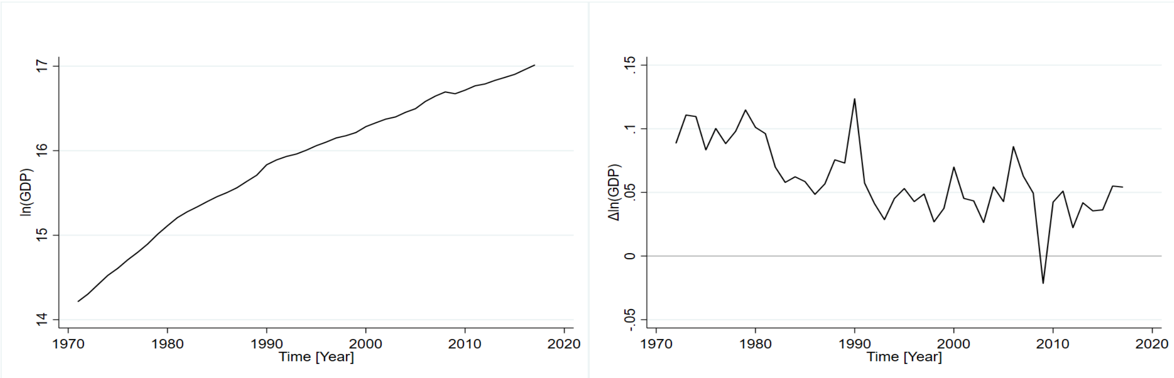


Figure 9: Level and First Difference Plots of GDP - OECD Europe. Source: OECD iLibrary

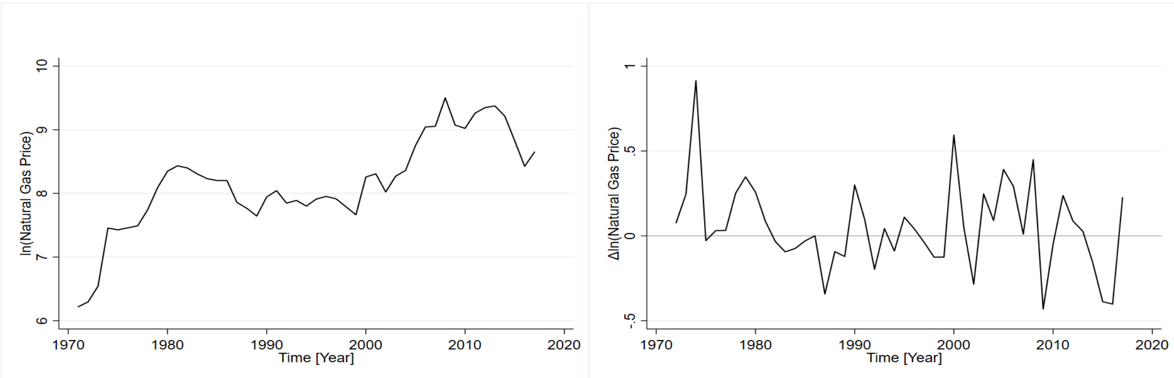


Figure 10: Level and First Difference Plots of Natural Gas Price - OECD Europe. Source: World Bank

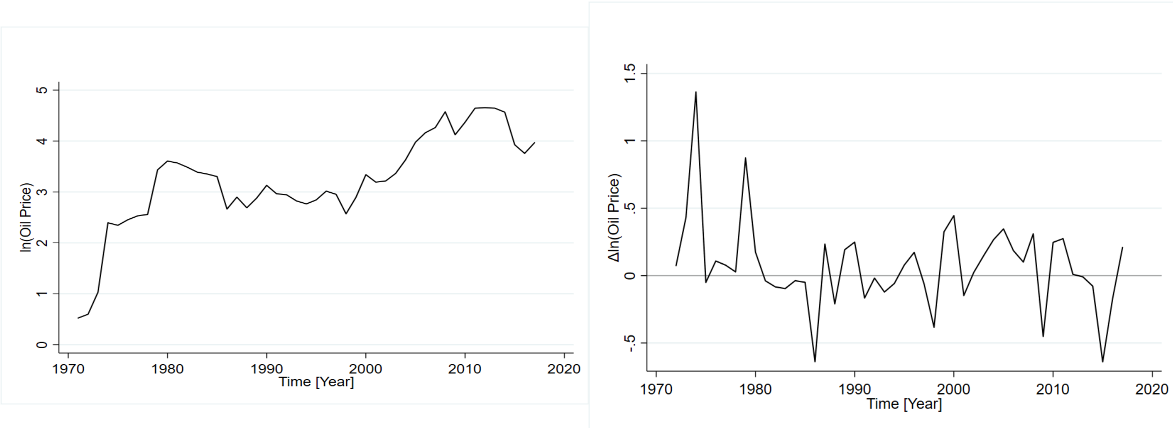


Figure 11: Level and First Difference Plots of Oil Price - OECD Europe. Source: World Bank

B.1.2 OECD Americas

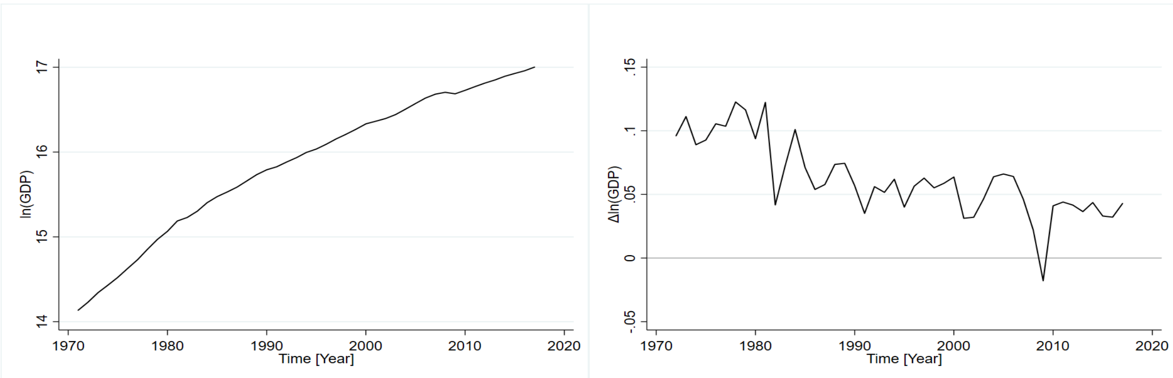


Figure 12: Level and First Difference Plots of GDP - OECD Americas. Source: OECD iLibrary

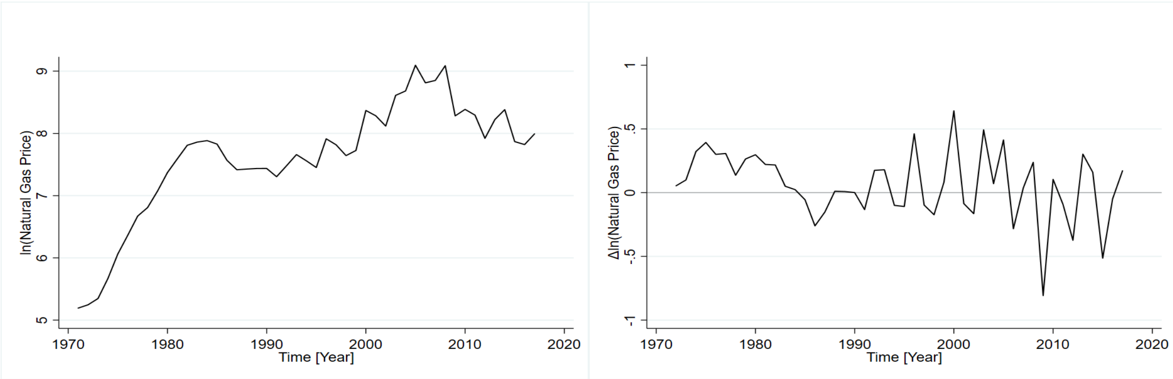


Figure 13: Level and First Difference Plots of Natural Gas Price - OECD Americas. Source: World Bank

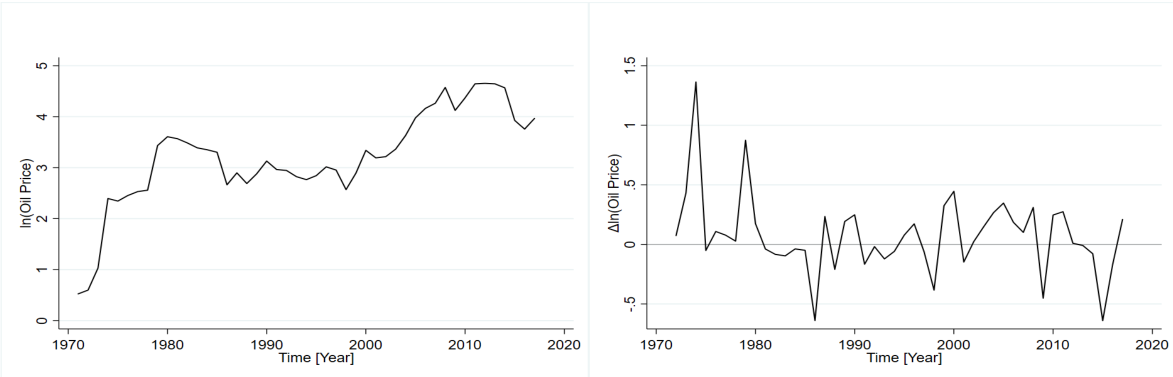


Figure 14: Level and First Difference Plots of Oil Price - OECD Americas. Source: World Bank

B.1.3 OECD Asia & Oceania

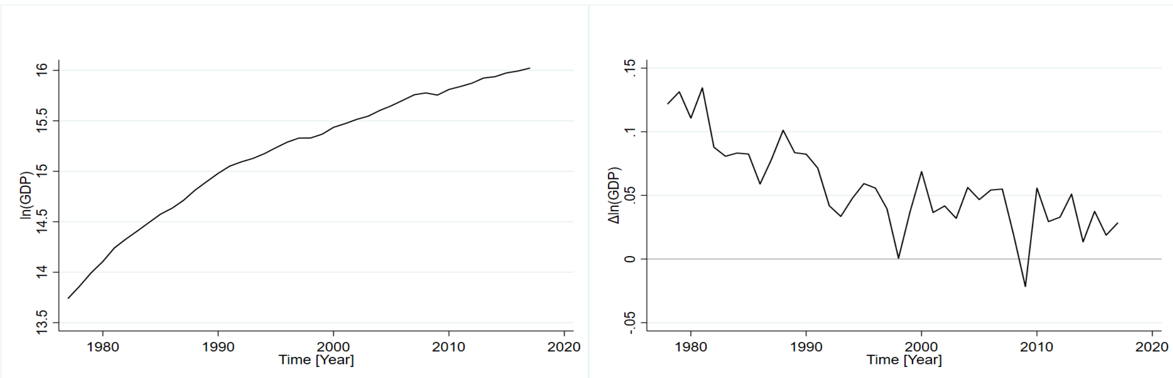


Figure 15: Level and First Difference Plots of GDP - OECD Asia & Oceania. Source: OECD iLibrary

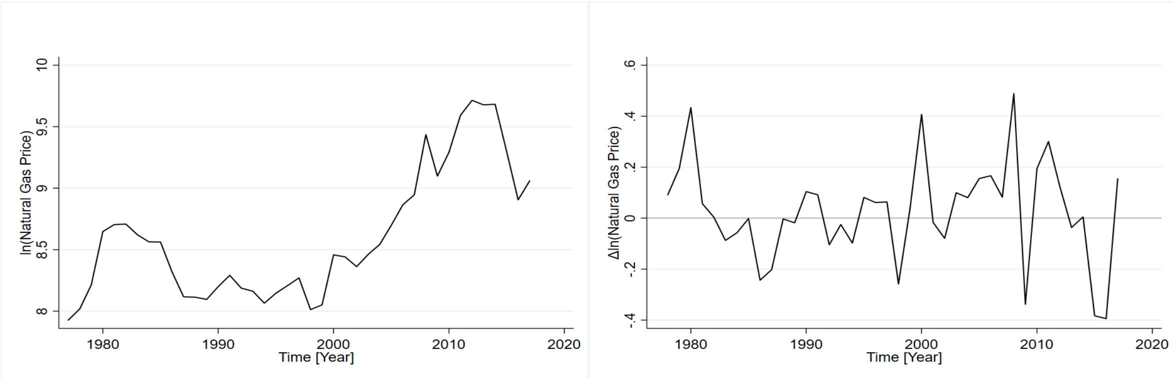


Figure 16: Level and First Difference Plots of Natural Gas Price - OECD Asia & Oceania. Source: World Bank

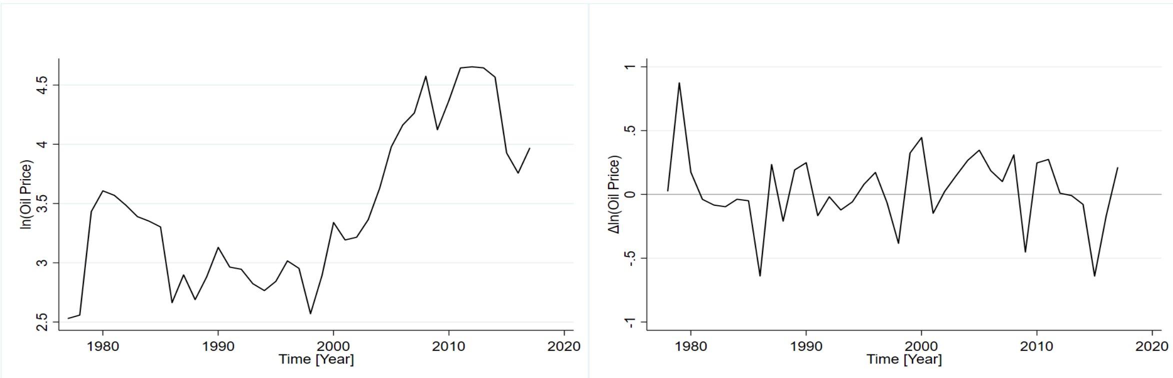
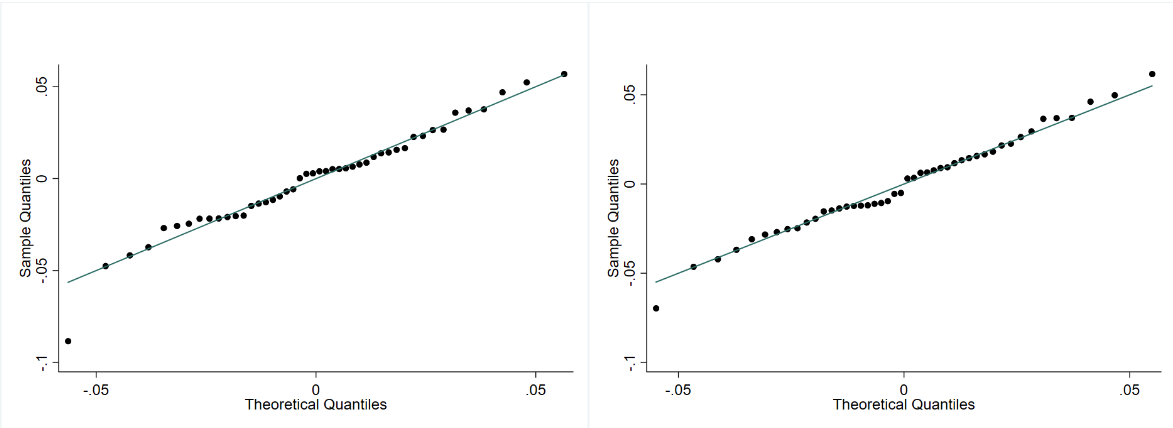


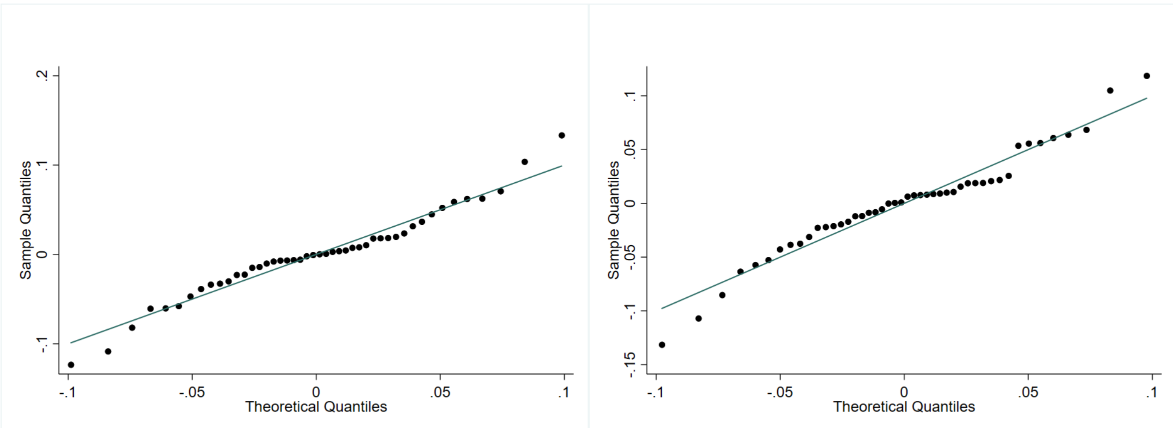
Figure 17: Level and First Difference Plots of Oil Price - OECD Asia & Oceania. Source: World Bank

B.2 Quantile-Quantile Plots



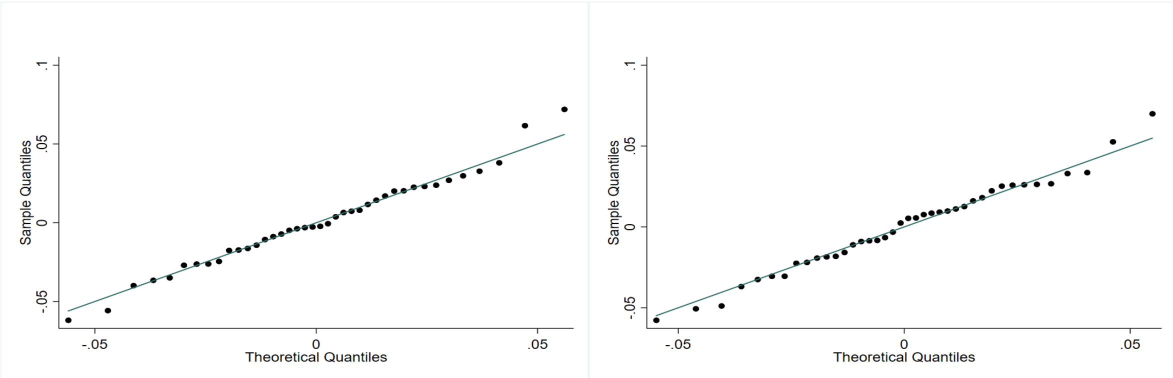
(a) ARDL(1, 0, 0, 1) **(b)** ARDL(1, 1, 1, 1)

Figure 18: Quantile-Quantile Plots for Residuals - Europe - Industry



(a) ARDL(1, 1, 0, 0) **(b)** ARDL(1, 1, 1, 1)

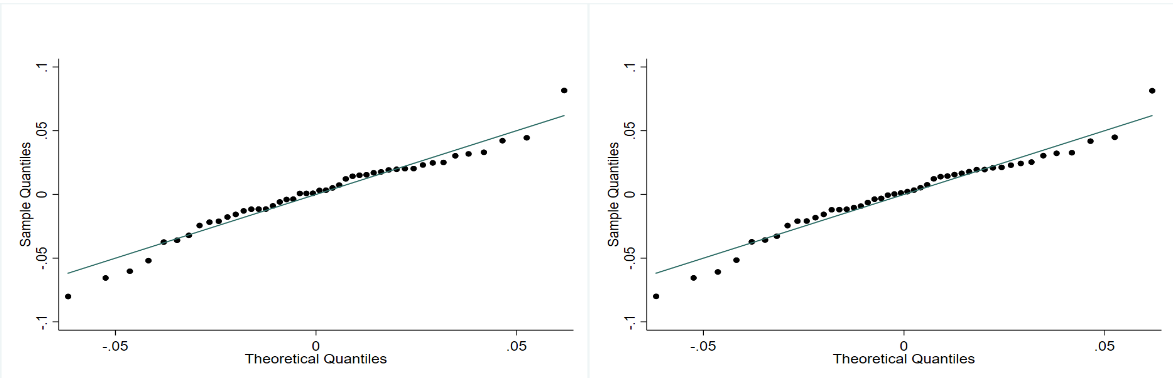
Figure 19: Quantile-Quantile Plots for Residuals - Americas - Industry



(a) ARDL(1, 0, 1, 0)

(b) ARDL(1, 1, 1, 1)

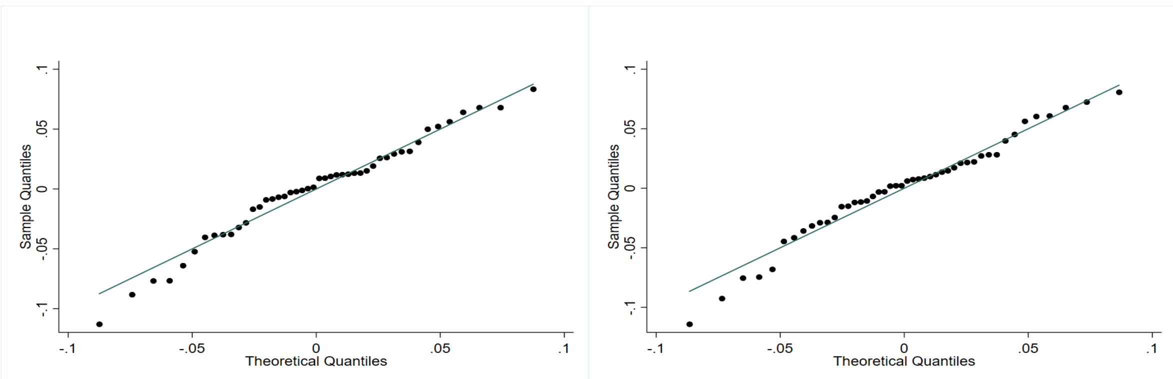
Figure 20: Quantile-Quantile Plots for Residuals - Asia & Oceania - Industry



(a) ARDL(1, 0, 0, 1)

(b) ARDL(1, 1, 1, 1)

Figure 21: Quantile-Quantile Plots for Residuals - Europe - Overall

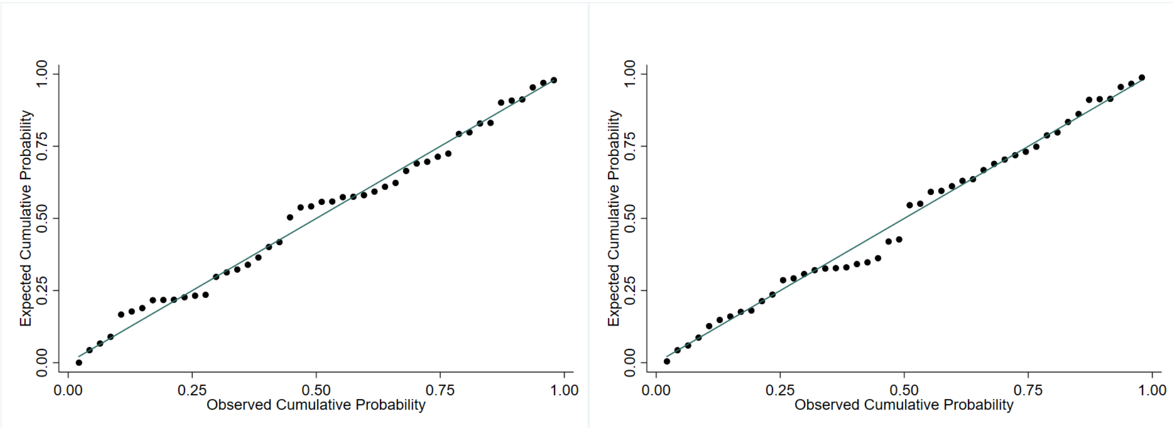


(a) ARDL(1, 0, 0, 1)

(b) ARDL(1, 1, 1, 1)

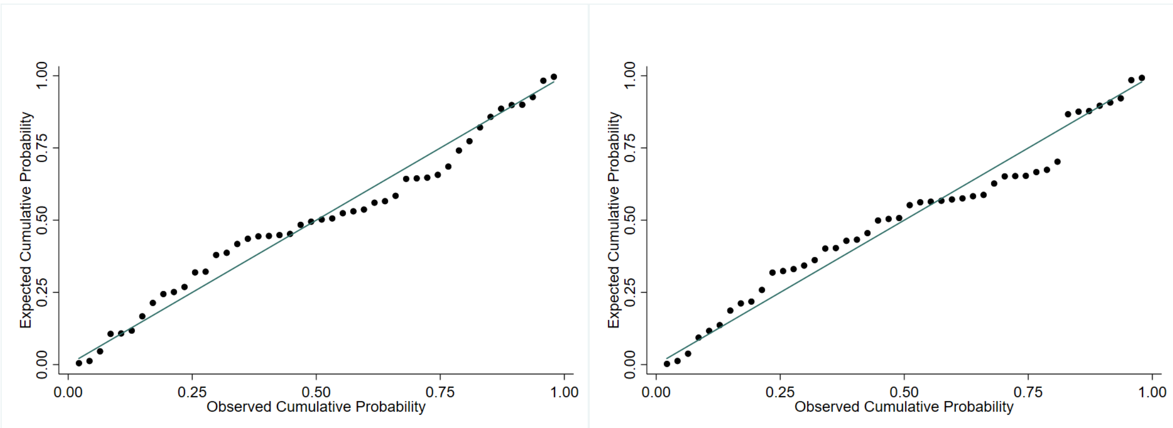
Figure 22: Quantile-Quantile Plots for Residuals - Europe - Sub-industry

B.3 Normal-Probability Plots



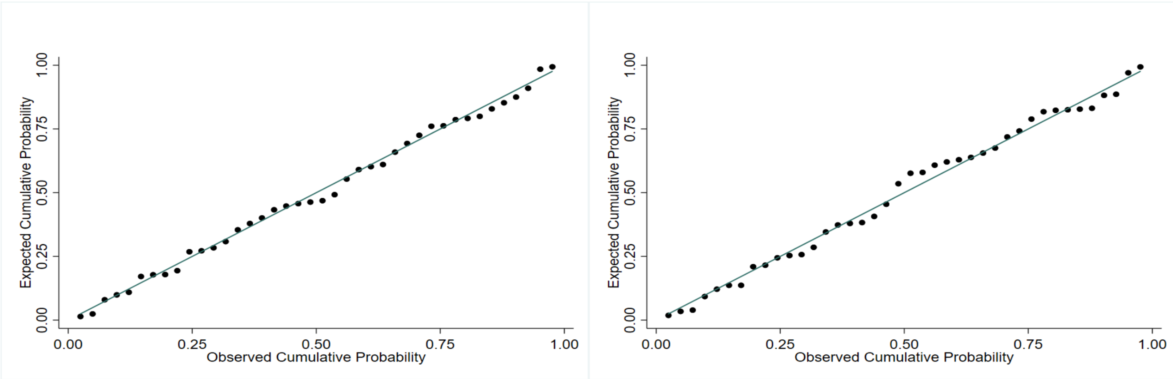
(a) ARDL(1, 0, 0, 1) **(b)** ARDL(1, 1, 1, 1)

Figure 23: Normal-Probability Plots for Residuals - Europe - Industry



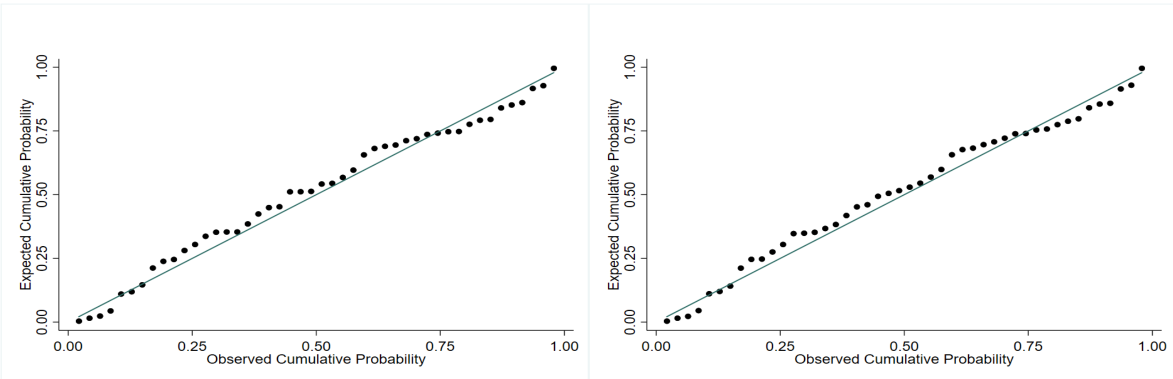
(a) ARDL(1, 1, 0, 0) **(b)** ARDL(1, 1, 1, 1)

Figure 24: Normal-Probability Plots for Residuals - Americas - Industry



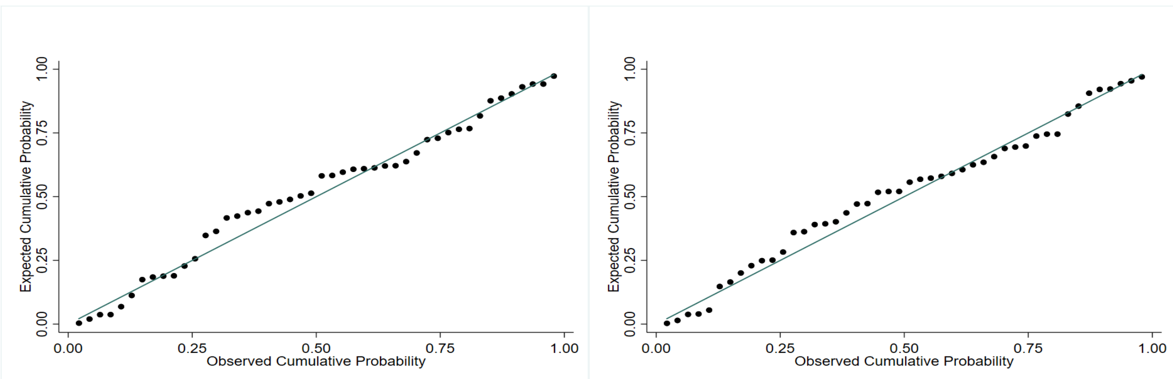
(a) ARDL(1, 0, 1, 0) **(b)** ARDL(1, 1, 1, 1)

Figure 25: Normal-Probability Plots for Residuals - Asia & Oceania - Industry



(a) ARDL(1, 0, 0, 1) **(b)** ARDL(1, 1, 1, 1)

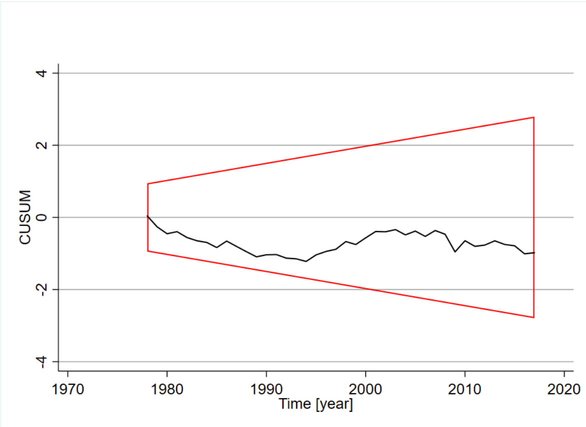
Figure 26: Normal-Probability Plots for Residuals - Europe - Overall



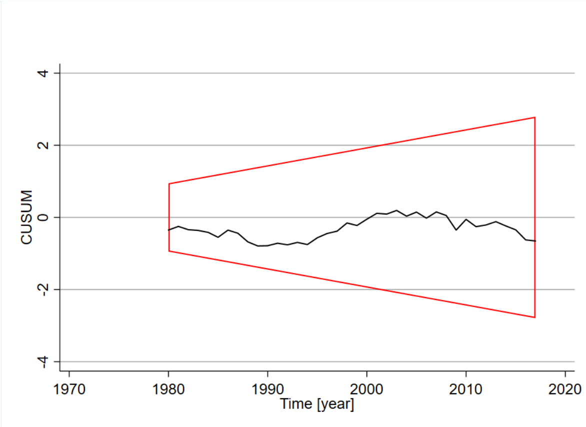
(a) ARDL(1, 0, 0, 1) **(b)** ARDL(1, 1, 1, 1)

Figure 27: Normal-Probability Plots for Residuals - Europe - Sub-industry

B.4 CUSUM Plots

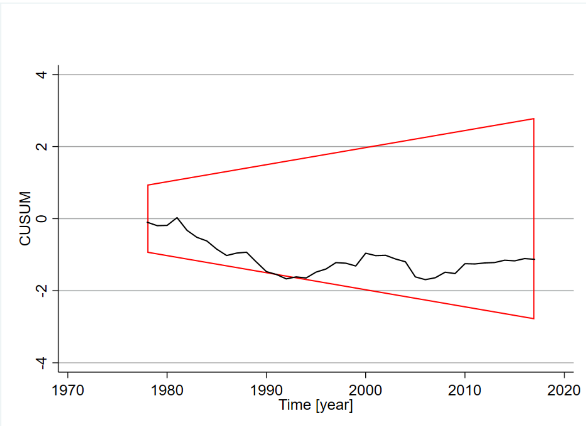


(a) ARDL(1, 0, 0, 1)

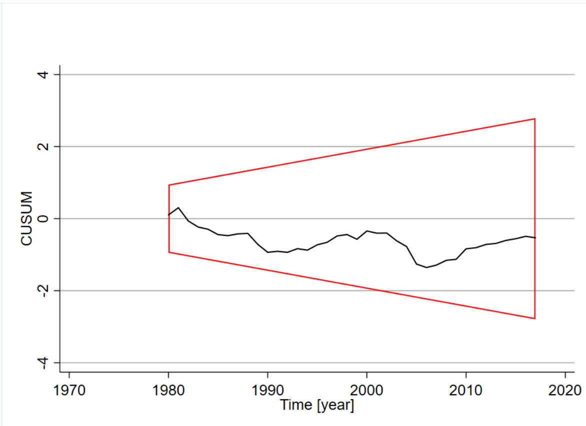


(b) ARDL(1, 1, 1, 1)

Figure 28: Recursive CUSUM Plots with 95 % Confidence Bound around the Null - Europe - Industry

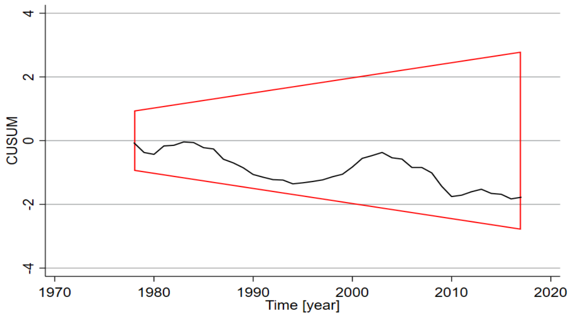


(a) ARDL(1, 1, 0, 0)

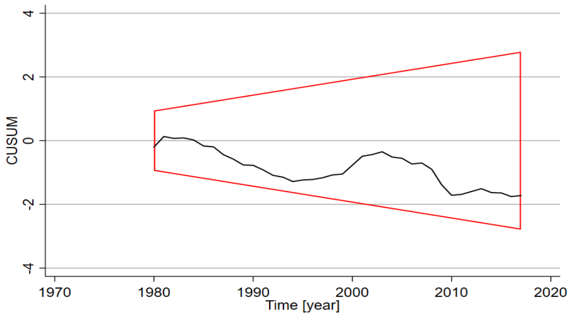


(b) ARDL(1, 1, 1, 1)

Figure 29: Recursive CUSUM Plots with 95 % Confidence Bound around the Null - Americas - Industry

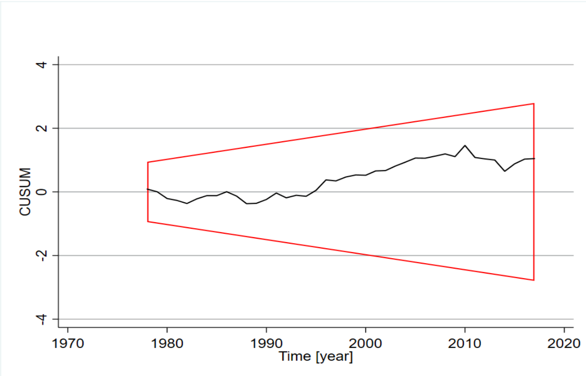


(a) ARDL(1, 0, 1, 0)

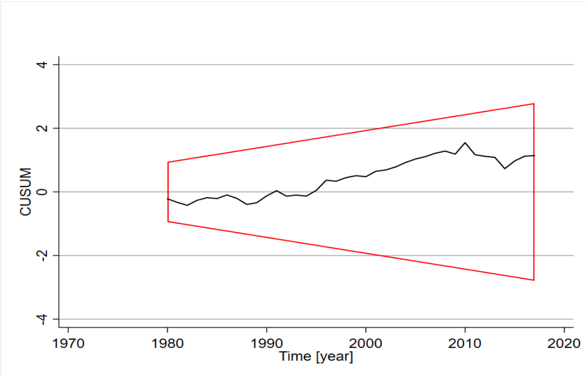


(b) ARDL(1, 1, 1, 1)

Figure 30: Recursive CUSUM Plots with 95 % Confidence Bound around the Null - Asia & Oceania - Industry

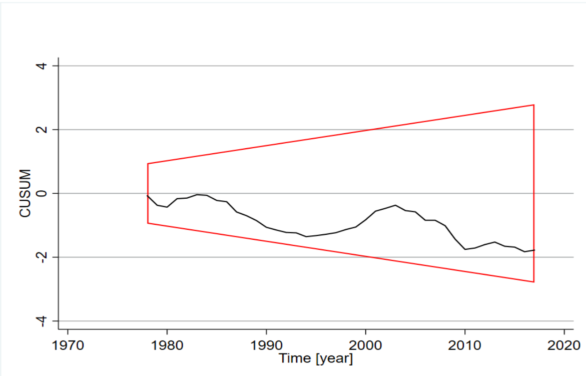


(a) ARDL(1, 0, 0, 1)

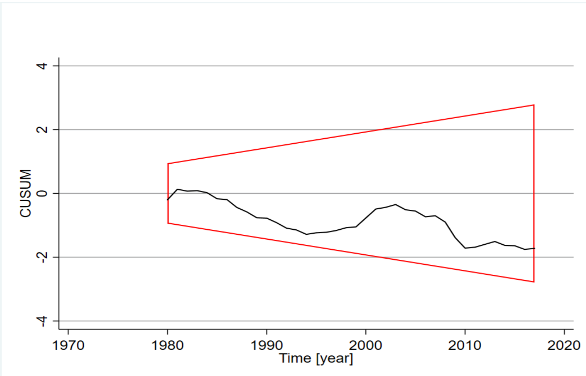


(b) ARDL(1, 1, 1, 1)

Figure 31: Recursive CUSUM Plots with 95 % Confidence Bound around the Null - Europe - Overall



(a) ARDL(1, 0, 0, 1)



(b) ARDL(1, 1, 1, 1)

Figure 32: Recursive CUSUM Plots with 95 % Confidence Bound around the Null - Europe - Sub-industry

C Appendix: Tables

C.1 Additional Tables - Model Results

C.1.1 Extended ARDL-results Tables

Model	**Europe - Industry**		
	ARDL(1, 0, 0, 1)		
Independent variable	Coefficient	Standard Error	p-value
$\ln(\text{Cons. of NG})_{t-1}$	0.731	0.046	0.000
$\ln(\text{GDP})_t$	0.075	0.013	0.000
$\ln(\text{GDP})_{t-1}$	-	-	-
$\ln(\text{NG Price})_t$	-0.043	0.037	0.251
$\ln(\text{NG Price})_{t-1}$	-	-	-
$\ln(\text{Oil Price})_t$	0.067	0.026	0.015
$\ln(\text{Oil Price})_{t-1}$	-0.076	0.016	0.000
Constant	2.295	0.477	0.000

Table 21: Extended ARDL-results - Europe - Industry - ARDL(1, 0, 0, 1)

Model	**Americas - Industry**		
	ARDL(1, 1, 0, 0)		
Independent variable	Coefficient	Standard Error	p-value
$\ln(\text{Cons. of NG})_{t-1}$	0.0558	0.101	0.000
$\ln(\text{GDP})_t$	2.191	0.458	0.000
$\ln(\text{GDP})_{t-1}$	-2.117	0.441	0.000
$\ln(\text{NG Price})_t$	-0.054	0.020	0.011
$\ln(\text{NG Price})_{t-1}$	-	-	-
$\ln(\text{Oil Price})_t$	0.002	0.017	0.906
$\ln(\text{Oil Price})_{t-1}$	-	-	-
Constant	4.457	1.390	0.003

Table 22: Extended ARDL-results - Americas - Industry - ARDL(1, 1, 0, 0)

Model	**Asia & Oceania - Industry**		
	ARDL(1, 0, 1, 0)		
Independent variable	Coefficient	Standard Error	p-value
$\ln(\text{Cons. of NG})_{t-1}$	0.755	0.076	0.000
$\ln(\text{GDP})_t$	0.209	0.070	0.005
$\ln(\text{GDP})_{t-1}$	-	-	-
$\ln(\text{NG Price})_t$	0.078	0.055	0.164
$\ln(\text{NG Price})_{t-1}$	-0.091	0.030	0.004
$\ln(\text{Oil Price})_t$	0.007	0.030	0.814
$\ln(\text{Oil Price})_{t-1}$	-	-	-
Constant	-0.680	0.417	0.112

Table 23: Extended ARDL-results - Asia & Oceania - Industry - ARDL(1, 0, 1, 0)

Model	Europe - Overall		
	ARDL(1, 0, 0, 1)		
Independent variable	Coefficient	Standard Error	p-value
ln(Cons. of NG)$_{t-1}$	0.828	0.051	0.000
ln(GDP)$_t$	0.093	0.025	0.001
ln(GDP)$_{t-1}$	-	-	-
ln(NG Price)$_t$	0.050	0.040	0.213
ln(NG Price)$_{t-1}$	-	-	-
ln(Oil Price)$_t$	-0.012	0.029	0.673
ln(Oil Price)$_{t-1}$	-0.074	0.017	0.000
Constant	0.612	0.391	0.126

Table 24: Extended ARDL-results - Europe - Overall - ARDL(1, 0, 0, 1)

Model	Europe - Sub-industry		
	ARDL(1, 0, 0, 1)		
Independent variable	Coefficient	Standard Error	p-value
ln(Cons. of NG)$_{t-1}$	0.483	0.094	0.000
ln(GDP)$_t$	0.044	0.015	0.007
ln(GDP)$_{t-1}$	-	-	-
ln(NG Price)$_t$	-0.087	0.056	0.130
ln(NG Price)$_{t-1}$	-	-	-
ln(Oil Price)$_t$	0.113	0.040	0.008
ln(Oil Price)$_{t-1}$	-0.068	0.024	0.008
Constant	5.058	0.963	0.000

Table 25: Extended ARDL-results - Europe - Sub-industry - ARDL(1, 0, 0, 1)

C.1.2 Extended ARDL-results Tables - ARDL(1, 1, 1, 1)

Model	**Europe - Industry**		
	ARDL(1, 1, 1, 1)		
Independent variable	Coefficient	Standard Error	p-value
ln(Cons. of NG)$_{t-1}$	0.720	0.051	0.000
ln(GDP)$_t$	0.322	0.287	0.269
ln(GDP)$_{t-1}$	-0.237	0.282	0.405
ln(NG Price)$_t$	-0.037	0.037	0.326
ln(NG Price)$_{t-1}$	-0.025	0.039	0.517
ln(Oil Price)$_t$	0.056	0.029	0.063
ln(Oil Price)$_{t-1}$	-0.052	0.030	0.090
Constant	2.356	0.673	0.001

Table 26: Extended ARDL-results - Europe - Industry - ARDL(1, 1, 1, 1)

Model	**Americas - Industry**		
	ARDL(1, 1, 1, 1)		
Independent variable	Coefficient	Standard Error	p-value
ln(Cons. of NG)$_{t-1}$	0.548	0.104	0.000
ln(GDP)$_t$	2.035	0.495	0.000
ln(GDP)$_{t-1}$	-1.961	0.480	0.000
ln(NG Price)$_t$	-0.040	0.038	0.304
ln(NG Price)$_{t-1}$	-0.015	0.039	0.700
ln(Oil Price)$_t$	0.017	0.030	0.570
ln(Oil Price)$_{t-1}$	-0.018	0.030	0.556
Constant	4.592	1.438	0.003

Table 27: Extended ARDL-results - Americas - Industry - ARDL(1, 1, 1, 1)

Model	**Asia & Oceania - Industry**		
	ARDL(1, 1, 1, 1)		
Independent variable	Coefficient	Standard Error	p-value
ln(Cons. of NG)$_{t-1}$	0.768	0.078	0.000
ln(GDP)$_t$	0.535	0.313	0.097
ln(GDP)$_{t-1}$	-0.322	0.302	0.294
ln(NG Price)$_t$	0.077	0.057	0.187
ln(NG Price)$_{t-1}$	-0.077	0.039	0.055
ln(Oil Price)$_t$	0.001	0.033	0.975
ln(Oil Price)$_{t-1}$	-0.006	0.037	0.864
Constant	-0.948	0.505	0.069

Table 28: Extended ARDL-results - Asia & Oceania - Industry - ARDL(1, 1, 1, 1)

Model	Europe - Overall		
	ARDL(1, 1, 1, 1)		
Independent variable	Coefficient	Standard Error	p-value
ln(Cons. of NG)$_{t-1}$	0.830	0.056	0.000
ln(GDP)$_t$	0.129	0.324	0.693
ln(GDP)$_{t-1}$	-0.036	0.322	0.911
ln(NG Price)$_t$	0.051	0.041	0.226
ln(NG Price)$_{t-1}$	0.002	0.041	0.952
ln(Oil Price)$_t$	-0.014	0.033	0.677
ln(Oil Price)$_{t-1}$	-0.074	0.032	0.025
Constant	0.568	0.583	0.336

Table 29: Extended ARDL-results - Europe - Overall - ARDL(1, 1, 1, 1)

Model	Europe - Sub-industry		
	ARDL(1, 1, 1, 1)		
Independent variable	Coefficient	Standard Error	p-value
ln(Cons. of NG)$_{t-1}$	0.451	0.105	0.000
ln(GDP)$_t$	-0.290	0.449	0.522
ln(GDP)$_{t-1}$	0.328	0.438	0.459
ln(NG Price)$_t$	-0.088	0.058	0.133
ln(NG Price)$_{t-1}$	-0.045	0.061	0.464
ln(Oil Price)$_t$	0.125	0.045	0.008
ln(Oil Price)$_{t-1}$	-0.043	0.048	0.373
Constant	5.755	1.276	0.000

Table 30: Extended ARDL-results - Europe - Sub-industry - ARDL(1, 1, 1, 1)

C.1.3 ARDL-results Tables - Comparisons

Model	Europe - Industry	
	ARDL(1, 0, 0, 1)	ARDL(1, 1, 1, 1)
Ind. var. / Dep. var.	ln(Cons. of NG)	ln(Cons. of NG)
ln(Cons. of NG)$_{t-1}$	0.731** (0.046)	0.720** (0.051)
ln(GDP)$_t$	0.075** (0.013)	0.322 (0.287)
ln(GDP)$_{t-1}$	-	-0.237 (0.282)
ln(NG Price)$_t$	-0.043 (0.037)	-0.037 (0.037)
ln(NG Price)$_{t-1}$	-	-0.025 (0.039)
ln(Oil Price)$_t$	0.067* (0.026)	0.056 (0.029)
ln(Oil Price)$_{t-1}$	-0.076** (0.016)	-0.052 (0.030)
Constant	2.295** (0.477)	2.356 (0.673)

Table 31: Comparison of ARDL-results from different specifications - Europe - Industry. The estimated coefficient is outside the parenthesis, and the corresponding standard error inside the parenthesis. (One asterisk indicates significance at 5 % level and two at 1 % level)

Model	Americas - Industry	
	ARDL(1, 1, 0, 0)	ARDL(1, 1, 1, 1)
Ind. var. / Dep. var.	ln(Cons. of NG)	ln(Cons. of NG)
ln(Cons. of NG)$_{t-1}$	0.558** (0.101)	0.548** (0.104)
ln(GDP)$_t$	2.191** (0.458)	2.035** (0.495)
ln(GDP)$_{t-1}$	-2.117** (0.441)	-1.961** (0.480)
ln(NG Price)$_t$	-0.054* (0.020)	-0.040 (0.038)
ln(NG Price)$_{t-1}$	-	-0.015 (0.039)
ln(Oil Price)$_t$	0.002 (0.017)	0.017 (0.030)
ln(Oil Price)$_{t-1}$	-	-0.018 (0.030)
Constant	4.457** (1.390)	4.592** (1.438)

Table 32: Comparison of ARDL-results from different specifications - Americas - Industry. The estimated coefficient is outside the parenthesis, and the corresponding standard error inside the parenthesis. (One asterisk indicates significance at 5 % level and two at 1 % level)

Model	Asia & Oceania - Industry	
	ARDL(1, 0, 1, 0)	ARDL(1, 1, 1, 1)
Ind. var. / Dep. var.	ln(Cons. of NG)	ln(Cons. of NG)
ln(Cons. of NG)$_{t-1}$	0.755** (0.076)	0.768** (0.078)
ln(GDP)$_t$	0.209** (0.070)	0.535 (0.313)
ln(GDP)$_{t-1}$	-	-0.322 (0.302)
ln(NG Price)$_t$	0.078 (0.055)	0.077 (0.057)
ln(NG Price)$_{t-1}$	-0.091** (0.030)	-0.077 (0.039)
ln(Oil Price)$_t$	0.007 (0.030)	0.001 (0.033)
ln(Oil Price)$_{t-1}$	-	-0.006 (0.037)
Constant	-0.680 (0.417)	-0.948 (0.505)

Table 33: Comparison of ARDL-results from different specifications - Asia & Oceania - Industry. The estimated coefficient is outside the parenthesis, and the corresponding standard error inside the parenthesis. (One asterisk indicates significance at 5 % level and two at 1 % level)

Model	Europe - Overall	
	ARDL(1, 0, 0, 1)	ARDL(1, 1, 1, 1)
Ind. var. / Dep. var.	ln(Cons. of NG)	ln(Cons. of NG)
ln(Cons. of NG)$_{t-1}$	0.828** (0.051)	0.830** (0.056)
ln(GDP)$_t$	0.093** (0.025)	0.129 (0.324)
ln(GDP)$_{t-1}$	-	-0.036 (0.322)
ln(NG Price)$_t$	0.050 (0.040)	0.051 (0.041)
ln(NG Price)$_{t-1}$	-	0.002 (0.041)
ln(Oil Price)$_t$	-0.012 (0.029)	-0.014 (0.033)
ln(Oil Price)$_{t-1}$	-0.074** (0.017)	-0.074* (0.032)
Constant	0.612 (0.391)	0.568 (0.583)

Table 34: Comparison of ARDL-results from different specifications - Europe - Overall. The estimated coefficient is outside the parenthesis, and the corresponding standard error inside the parenthesis. (One asterisk indicates significance at 5 % level and two at 1 % level)

Model	Europe - Sub-industry	
	ARDL(1, 0, 0, 1)	ARDL(1, 1, 1, 1)
Ind. var. / Dep. var.	ln(Cons. of NG)	ln(Cons. of NG)
ln(Cons. of NG)$_{t-1}$	0.483** (0.094)	0.451** (0.105)
ln(GDP)$_t$	0.044** (0.015)	-0.290 (0.449)
ln(GDP)$_{t-1}$	-	0.328 (0.438)
ln(NG Price)$_t$	-0.087 (0.056)	-0.088 (0.058)
ln(NG Price)$_{t-1}$	-	-0.045 (0.061)
ln(Oil Price)$_t$	0.113** (0.040)	0.125** (0.045)
ln(Oil Price)$_{t-1}$	-0.068** (0.024)	-0.043 (0.048)
Constant	5.058** (0.963)	5.755** (1.276)

Table 35: Comparison of ARDL-results from different specifications - Europe - Sub-industry. The estimated coefficient is outside the parenthesis, and the corresponding standard error inside the parenthesis. (One asterisk indicates significance at 5 % level and two at 1 % level)

C.1.4 Extended EC-results Tables

Model	**Europe - Industry**		
	ARDL(1, 0, 0, 1)		
Independent Variable	Coefficient	Standard Error	p-value
Speed of Adjustment	-0.269	0.046	0.000
Long-run			
Independent Variable	Coefficient	Standard Error	p-value
ln(GDP)	0.277	0.038	0.000
ln(NG Price)	-0.158	0.135	0.247
ln(Oil Price)	-0.034	0.112	0.762
Short-run			
Independent Variable	Coefficient	Standard Error	p-value
Δln(GDP)$_t$	-	-	-
Δln(NG Price)$_t$	-	-	-
Δln(Oil Price)$_t$	0.077	0.016	0.000
Constant	2.295	0.477	0.000

Table 36: Extended EC-results - Europe - Industry - ARDL(1, 0, 0, 1)

Model	**Americas - Industry**		
	ARDL(1, 1, 0, 0)		
Independent Variable	Coefficient	Standard Error	p-value
Speed of Adjustment	-0.442	0.101	0.000
Long-run			
Independent Variable	Coefficient	Standard Error	p-value
ln(GDP)	0.167	0.076	0.034
ln(NG Price)	-0.123	0.047	0.012
ln(Oil Price)	0.005	0.038	0.905
Short-run			
Independent Variable	Coefficient	Standard Error	p-value
Δln(GDP)$_t$	2.117	0.441	0.000
Δln(NG Price)$_t$	-	-	-
Δln(Oil Price)$_t$	-	-	-
Constant	4.457	1.390	0.003

Table 37: Extended EC-results - Americas - Industry - ARDL(1, 1, 0, 0)

Model	**Europe - Sub-industry**		
	ARDL(1, 0, 0, 1)		
Independent Variable	Coefficient	Standard Error	p-value
Speed of Adjustment	-0.517	0.094	0.000
Long-run			
Independent Variable	Coefficient	Standard Error	p-value
ln(GDP)	0.085	0.029	0.006
ln(NG Price)	-0.168	0.112	0.141
ln(Oil Price)	0.087	0.089	0.336
Short-run			
Independent Variable	Coefficient	Standard Error	p-value
Δln(GDP)$_t$	-	-	-
Δln(NG Price)$_t$	-	-	-
Δln(Oil Price)$_t$	0.068	0.024	0.008
Constant	5.058	0.963	0.000

Table 38: Extended EC-results - Europe - Sub-industry - ARDL(1, 0, 0, 1)

C.1.5 Extended EC-results Tables - ARDL(1, 1, 1, 1)

Model	**Europe - Industry**		
	ARDL(1, 1, 1, 1)		
Independent Variable	Coefficient	Standard Error	p-value
Speed of Adjustment	-0.280	0.051	0.000
Long-run			
Independent Variable	Coefficient	Standard Error	p-value
ln(GDP)	0.303	0.049	0.000
ln(NG Price)	-0.222	0.180	0.224
ln(Oil Price)	0.014	0.153	0.928
Short-run			
Independent Variable	Coefficient	Standard Error	p-value
$\Delta\ln(\text{GDP})_t$	0.237	0.282	0.405
$\Delta\ln(\text{NG Price})_t$	0.025	0.039	0.517
$\Delta\ln(\text{Oil Price})_t$	0.052	0.030	0.090
Constant	2.356	0.673	0.001

Table 39: Extended EC-results - Europe - Industry - ARDL(1, 1, 1, 1)

Model	**Americas - Industry**		
	ARDL(1, 1, 1, 1)		
Independent Variable	Coefficient	Standard Error	p-value
Speed of Adjustment	-0.452	0.104	0.000
Long-run			
Independent Variable	Coefficient	Standard Error	p-value
ln(GDP)	0.164	0.075	0.035
ln(NG Price)	-0.122	0.048	0.016
ln(Oil Price)	-0.0004	0.039	0.992
Short-run			
Independent Variable	Coefficient	Standard Error	p-value
$\Delta\ln(\text{GDP})_t$	1.961	0.480	0.000
$\Delta\ln(\text{NG Price})_t$	0.015	0.039	0.700
$\Delta\ln(\text{Oil Price})_t$	0.018	0.030	0.556
Constant	4.592	1.438	0.003

Table 40: Extended EC-results - Americas - Industry - ARDL(1, 1, 1, 1)

Model	Europe - Sub-industry		
	ARDL(1, 1, 1, 1)		
Independent Variable	Coefficient	Standard Error	p-value
Speed of Adjustment	-0.549	0.105	0.000
Long-run			
Independent Variable	Coefficient	Standard Error	p-value
ln(GDP)	0.069	0.035	0.057
ln(NG Price)	-0.243	0.141	0.094
ln(Oil Price)	0.149	0.113	0.194
Short-run			
Independent Variable	Coefficient	Standard Error	p-value
Δln(GDP)$_t$	-0.328	0.438	0.459
Δln(NG Price)$_t$	0.045	0.061	0.464
Δln(Oil Price)$_t$	0.043	0.048	0.373
Constant	5.755	1.276	0.000

Table 41: Extended EC-results - Europe - Sub-industry - ARDL(1, 1, 1, 1)

C.1.6 EC-results Tables - Comparisons

Model	**Europe - Industry**	
	ARDL(1, 0, 0, 1)	ARDL(1, 1, 1, 1)
Ind. var. / Dep. var.	Δln(Cons. of NG)	Δln(Cons. of NG)
Speed of Adjustment	-0.269** (0.046)	-0.280** (0.051)
Long-run		
ln(GDP)	0.0277** (0.038)	0.303** (0.049)
ln(NG Price)	-0.158 (0.135)	-0.222 (0.180)
ln(Oil Price)	-0.034 (0.112)	0.014 (0.153)
Short-run		
$\Delta\ln(\text{GDP})_t$	-	0.237 (0.282)
$\Delta\ln(\text{NG Price})_t$	-	0.025 (0.039)
$\Delta\ln(\text{Oil Price})_t$	0.076** (0.016)	0.052 (0.030)
Constant	2.295** (0.477)	2.356** (0.673)

Table 42: Comparison of EC-results from different specifications - Europe - Industry. The estimated coefficient is outside the parenthesis, and the corresponding standard error inside the parenthesis. (One asterisk indicates significance at 5 % level and two at 1 % level))

Model	**Americas - Industry**	
	ARDL(1, 1, 0, 0)	ARDL(1, 1, 1, 1)
Ind. var. / Dep. var.	Δln(Cons. of NG)	Δln(Cons. of NG)
Speed of Adjustment	-0.442** (0.101)	-0.452** (0.104)
Long-run		
ln(GDP)	0.167* (0.076)	0.164* (0.075)
ln(NG Price)	-0.123* (0.047)	-0.122* (0.048)
ln(Oil Price)	0.005 (0.038)	-0.0004 (0.039)
Short-run		
$\Delta\ln(\text{GDP})_t$	2.117** (0.441)	1.961** (0.480)
$\Delta\ln(\text{NG Price})_t$	-	0.015 (0.039)
$\Delta\ln(\text{Oil Price})_t$	-	0.018 (0.030)
Constant	4.457** (1.390)	4.592** (1.438)

Table 43: Comparison of EC-results from different specifications - Americas - Industry. The estimated coefficient is outside the parenthesis, and the corresponding standard error inside the parenthesis. (One asterisk indicates significance at 5 % level and two at 1 % level)

Model	**Europe - Sub-industry**	
	ARDL(1, 0, 0, 1)	ARDL(1, 1, 1, 1)
Ind. var. / Dep. var.	Δln(Cons. of NG)	Δln(Cons. of NG)
Speed of Adjustment	-0.517** (0.094)	-0.549** (0.105)
Long-run		
ln(GDP)	0.085** (0.029)	0.069 (0.035)
ln(NG Price)	-0.168 (0.112)	-0.243 (0.141)
ln(Oil Price)	0.087 (0.089)	0.149 (0.113)
Short-run		
Δln(GDP)$_t$	-	-0.328 (0.438)
Δln(NG Price)$_t$	-	0.045 (0.061)
Δln(Oil Price)$_t$	0.068** (0.024)	0.043 (0.048)
Constant	5.058** (0.963)	5.755** (1.276)

Table 44: Comparison of EC-results from different specifications - Europe - Sub-industry. The estimated coefficient is outside the parenthesis, and the corresponding standard error inside the parenthesis. (One asterisk indicates significance at 5 % level and two at 1 % level)

C.2 Cointegration - Tables for the ARDL(1, 1, 1, 1)-models

	F-Statistic	t-Statistic	Finite Sample	Asymptotic
Europe - Industry	17.738	-5.435	Yes**	Yes**
Americas - Industry	9.336	-4.330	Yes*	Yes**
Asia & Oceania - Industry	3.268	-2.984	No	No
Europe - Overall	15.639	-3.023	No	No
Europe - Sub-industry	8.112	-5.231	Yes**	Yes**

Table 45: Test-statistics for the Pesaran et al. (2001) Bounds Test - Case 3 - ARDL(1, 1, 1, 1) and Evaluation using Kripfganz & Schneider (2018) Critical Values. (One asterisk indicates significance at 5 % level, and two asterisks at 1 % level.)

Finite Sample - F-Statistic	**10 %**	**5 %**	**1 %**
Europe - Industry	[2.855 , 4.006]	[3.461 , 4.750]	[4.873 , 6.457]
Americas - Industry	[2.855 , 4.006]	[3.461 , 4.750]	[4.873 , 6.457]
Asia & Oceania - Industry	[2.888 , 4.066]	[3.521 , 4.851]	[5.015 , 6.684]
Europe - Overall	[2.855 , 4.006]	[3.461 , 4.750]	[4.873 , 6.457]
Europe - Sub-industry	[2.855 , 4.006]	[3.461 , 4.750]	[4.873 , 6.457]
Finite Sample - t-Statistic	**10 %**	**5 %**	**1 %**
Europe - Industry	[-2.550 , -3.438]	[-2.888 , -3.817]	[-3.567 , -4.569]
Americas - Industry	[-2.550 , -3.438]	[-2.888 , -3.817]	[-3.567 , -4.569]
Asia & Oceania - Industry	[-2.549 , -3.441]	[-2.896 , -3.833]	[-3.599 , -4.619]
Europe - Overall	[-2.550 , -3.438]	[-2.888 , -3.817]	[-3.567 , -4.569]
Europe - Sub-industry	[-2.550 , -3.438]	[-2.888 , -3.817]	[-3.567 , -4.569]
	10 %	**5 %**	**1 %**
Asymptotic - F-Statistic	[2.729 , 3.747]	[3.226 , 4.321]	[4.296 , 5.535]
	10 %	**5 %**	**1 %**
Asymptotic - t-Statistic	[-2.569 , -3.426]	[-2.864 , -3.744]	[-3.434 , -4.342]

Table 46: Kripfganz & Schneider (2018) Critical Values for the Pesaran et al. (2001) Bounds Test - Case 3 - ARDL(1, 1, 1, 1)

C.3 Diagnostics - Tables for the ARDL(1, 1, 1, 1)-models

Statistic	**Europe - Industry**	**Americas - Industry**	**Asia & Oceania - Industry**	**Europe - Overall**	**Europe - Sub-industry**
Serial Correlation - (Godfrey, 1978)	1.827 (0.1765)	2.486 (0.1149)	0.295 (0.5869)	0.743 (0.3887)	1.420 (0.2334)
Heteroskedasticity - (White, 1980)	44.59 (0.1285)	41.70 (0.2024)	37.45 (0.3574)	41.87 (0.1974)	36.76 (0.3875)
Model Misspecification - (Ramsey, 1969)	0.19 (0.9004)	0.62 (0.6069)	2.51 (0.0786)	0.49 (0.6899)	1.49 (0.2346)
Normality - (D'Agostino et al., 1990)	0.09 (0.9558)	2.59 (0.2735)	0.11 (0.9466)	2.92 (0.2317)	2.67 (0.2627)

Table 47: Diagnostic Tests for the ARDL-models - ARDL(1, 1, 1, 1). The test statistic is shown outside the parenthesis, and the corresponding p-value inside the parenthesis.

Statistic	**Europe - Industry**	**Americas - Industry**	**Asia & Oceania - Industry**	**Europe - Overall**	**Europe - Sub-industry**
Stability (Brown et al., 1975)	0.5178	0.5906	0.3893	0.5883	0.7168

Table 48: Brown et al. (1975) CUSUM test for the ARDL-models - ARDL(1, 1, 1, 1).

References

Ackah, I. (2014, September). Determinants of natural gas demand in Ghana. *OPEC Energu Review*, *38*(3), 272–295. Retrieved 2019-12-09, from https://onlinelibrary.wiley.com/doi/full/10.1111/opec.12026

Andersen, T. B., Nilsen, O. B., & Tveteras, R. (2011, September). How is demand for natural gas determined across European industrial sectors? *Energy Policy*, *39*(9), 5499–5508. Retrieved 2019-12-09, from https://www.sciencedirect.com/science/article/pii/S0301421511003843

BP. (2019). *Bp statistical review of world energy 2019* (Tech. Rep.). Author. Retrieved 2019-10-20, from https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2019-full-report.pdf

Brooks, C. (2014). *Introductory Econometrics for Finance* (3rd ed.). Cambridge University Press.

Brown, R. L., Durbin, J., & Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B (Methodological)*, *37*(2), 149–192.

Burke, P. J., & Yang, H. (2016, September). The price and income elasticities of natural gas demand: International evidence. *Energy Economics*, *59*, 466–474. Retrieved 2019-12-09, from https://www.sciencedirect.com/science/article/pii/S0140988316302420

Cuddington, J. T., & Dagher, L. (2015, January). Estimating Short and Long-Run Demand Elasticities: A Primer with Energy-Sector Applications. *The Energy Journal*, *36*(1), 185–209.

D'Agostino, R. B., Belanger, A., & D'Agostino Jr., R. B. (1990). A suggestion for using powerful and informative tests of normality. *The American Statistician*, *44*(4), 316-321. doi: 10.1080/00031305.1990.10475751

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association*, *74*(366), 427–431. doi: 10.2307/2286348

Dilaver, , Dilaver, Z., & Hunt, L. C. (2014, July). What drives natural gas consumption in Europe? Analysis and projections. *Journal of Natural Gas Science and Engineering*, *19*, 125–136. Retrieved 2019-10-12, from https://www.sciencedirect.com/science/article/pii/S1875510014000870

EIA, U. (n.d.). *What are the major factors affecting natural gas prices?* Retrieved 2019-12-15, from https://www.eia.gov/tools/faqs/faq.php?id=43&t=8

Erdogdu, E. (2010, January). Natural gas demand in Turkey. *Applied Energy*, *87*(1), 211–219. Retrieved 2019-12-09, from https://www.sciencedirect.com/science/article/pii/S0306261909002931

European Commision. (2015). *EU ETS Handbook.* Retrieved from https://ec.europa.eu/clima/sites/clima/files/docs/ets_handbook_en.pdf

Fuller, W. A. (1996). *Introduction to Statistical Time Series* (2nd ed.). New York: Wiley.

Gautam, T. K., & Paudel, K. P. (2018, September). The demand for natural gas in the Northeastern United States. *Energy*, *158*, 890–898. Retrieved 2019-12-09, from https://www.sciencedirect.com/science/article/pii/S0360544218311575#bib4

Godfrey, L. (1978). Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica*, *46*(6), 1293–1301.

Harold, J., Lyons, S., & Cullinan, J. (2015, September). The determinants of residential gas demand in Ireland. *Energy Economics*, *51*, 475–483. Retrieved 2019-12-09, from https://www.sciencedirect.com/science/article/pii/S0140988315002388

Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence-Based Nursing*, *18*(3), 66–67. doi: 10.1136/eb-2015-102129

IEA. (2018). *Natural gas information 2018* (Tech. Rep.). International Energy Agency.

IEA. (2019a). *Gas 2019 - Analysis and forecast to 2024* (Tech. Rep.). International Energy Agency.

IEA. (2019b). *Transforming industry through ccus* (Tech. Rep.). Paris: International Energy Agency.

Kripfganz, S., & Schneider, D. C. (2018, September). Estimating autoregressive distributed lag and equilibrium correction models. London.

Mills, A., Durepos, G., & Wiebe, E. (2010). *Encyclopedia of Case Study Research* (Vol. 1-0). Thousand Oaks, California: SAGE Publications, Inc. doi: 10.4135/9781412957397

Pesaran, M. H., & Shin, Y. (1995). *Long-Run Structural Modelling* (Cambridge Working Papers in Economics). Faculty of Economics, University of Cambridge.

Pesaran, M. H., & Shin, Y. (1999). An Autoregressive Distributed Lag Modelling Approach to Cointegration Analysis. *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium*.

Pesaran, M. H., Shin, Y., & Smith, R. J. (2001). Bounds Testing Approaches to the Analysis of Level Relationships. *Journal of Applied Econometrics*, *16*(3), 289–326.

Ramberg, D. J., & Parsons, J. E. (2012). The Weak Tie Between Natural Gas and Oil Prices. *The Energy Journal*, *33*(2). Retrieved from https://econpapers.repec.org/article/aenjournl/33-2-02.htm

Ramsey, J. B. (1969). Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, *31*(2), 350–371.

Raus, L. (2014, December). Methods of Natural Gas Transport. *Egypt Oil & Gas*. Retrieved 2019-10-20, from https://egyptoil-gas.com/features/methods-of-natural-gas-transport/

Wadud, Z., Dey, H. S., Kabir, A., & Khan, S. I. (2011, November). Modeling and forecasting natural gas demand in Bangladesh. *Energy Policy*, *39*(11), 7372–7380. Retrieved 2019-12-09, from https://www.sciencedirect.com/science/article/pii/S0301421511006872

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*(4), 817–838.

Woolridge, J. M. (2009). *Introductory Econometrics: A Modern Approach* (4th ed.). South-Western Cengage Learning.

Yu, Y., Zheng, X., & Han, Y. (2014, July). On the demand for natural gas in urban China. *Energy Policy*, *70*, 57–63. Retrieved 2019-12-09, from https://www.sciencedirect.com/science/article/pii/S0301421514001797

Zaretskaya, V., & Bradley, S. (2015, September). *Natural gas prices in Asia mainly linked to crude oil, but use of spot indexes increases.* Retrieved 2019-12-15, from https://www.eia.gov/todayinenergy/detail.php?id=23132

Zhang, Y., Ji, Q., & Fan, Y. (2018, February). The price and income elasticity of China's natural gas demand: A multi-sectoral perspective. *Energy Policy*, *113*, 332–341. Retrieved 2019-12-09, from https://www.sciencedirect.com/science/article/pii/S0301421517307516

Emilie C. Løberg

What determines the level of natural gas demand?

NTNU
Norwegian University of
Science and Technology