DISSERTATION

Econometric Modeling of Natural Gas Prices

Different Modeling Approaches and Case Studies related to Gas Markets

Written for the purpose of obtaining the academic degree

DOKTOR DER TECHNISCHEN WISSENSCHAFTEN (DR. TECHN.)

Submitted at

TECHNISCHE UNIVERSITAT WIEN

FACULTY OF ELECTRICAL ENGINEERING AND INFORMATION TECHNOLOGY

Under the supervision of

PD Dr. Dipl.-Ing. Johann Auer, Energy Economics Group E370-3

And reviewed & examined by

Professor Bacel Maddah, Department of Industrial Engineering and Management, American University of Beirut, AUB

Professor Anne Neumann, Department of Industrial Economics and Technology Management, Norwegian University of Science and Technology, NTNU

By

MSc. Hassan HAMIE

Mat. Nr. 1426379

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources/ resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Beirut, ______________________ ______________________

Date Signature

Acknowledgment

Firstly, I would like to express my sincere gratitude to my advisor Prof. Dr. Hans Auer for the continuous support through the time of my Ph.D thesis. I would like to thanks him for his support in my study and related research, for his patience, motivation, and humble character. Although, I was only able to be at the Energy Economics Group (EEG), Vienna for short and limited times, his guidance helped me in all the time of research and writing of this thesis. The time invested in his Ph.D students and fellows is appreciated and recognizable.

Besides my advisor, a wide range of people working for professional organization, provided me an opportunity to access confidential data, and share feedback and point of view on some of my analysis. To name a few of these organization: The European Energy Exchange, the government regulator for electricity and natural gas markets in Austria, E-Control, the Austrian Gas Grid Management AG, AGGM, etc.

I want to thank the members of the thesis defense committee for their insightful remarks. The constructive comments have helped me greatly in preparing for the final version of my thesis. Last but not the least, I would like to thank my family: my parents, my brother and sister for supporting me spiritually throughout writing this thesis. No words can describe how blessed I am, having such a family!

Abstract

There are many reasons why a better understanding of the stochastic process driving the prices of natural gas would be useful. This understanding would be helpful on the microeconomic level: providing an efficient tool for better forecasting the gas prices; aiding with the decision and timing of investment, as well as understanding the main variables that affect the wholesale prices of natural gas.

The question of whether liberalization of the gas industry has led to more competitive markets has attracted much interest among the scientific community. Classical mathematical regression tools, statistical tests, and optimization equilibrium problems, more precisely non-linear complementarity problems, were used to model gas markets and its effect on prices. On the basis of four challenges/ case studies, the aim of the thesis is to introduce these important methods widely used in econometrics, by linking them to a common "use case" in the subject of commodity pricing, such as natural gas.

First, the **Records theory** is used to study the effect of extreme gas prices and the probability of future records. The aim is to test the stability of three different regional gas markets (U.S, Europe and Asia), each having its own supply and demand characteristics. The $classical$ model is used for the case where gas prices are independent and identically distributed ($i.i.d\ case$). Alternative models, such as Yang model and the discrete-time random walk, are used, where the number of records grows faster than in the $i.i.d$ case and where records are not only concentrated among the first observations. The findings suggest that out of the three main regional gas markets, the Asian market seems to be less stable than the others, and that the probability of having a record in the coming years is the highest. The main advantage of such a model that has not been used previously for commodity markets, is that in spite of the non-independent and non-identically distributed properties of the data, the results are distribution free. Consequently, the applicant will not be concerned by identifying the distribution type and the complexity of the models is reduced.

Second, **Game theory** is used to test the concentration and behavior of gas suppliers in two different regional European gas markets, Austria and the Netherlands, each one of them represents a different evolutionary stage in the process of wholesale natural gas markets liberalization. The parametric method takes into account the classical Nash-Cournot equilibrium test, with assumptions on cost and demand functions, while the non-parametric method does not make any prior parametric assumptions, and thus allow greater freedom in modeling. This model is effective and validates the fact that suppliers in both markets follow the 'profit maximization' behavior even though one market is more liquid than the other. Interestingly, the findings also suggest that some of the gas suppliers maximize their 'utility function' not only by seeking profit but also by pursuing non-profit objectives, such as cooperative collusive behavior. Additionally, by using a non-parametric method other than the widely used classical time series methods, the gas market integration in Europe is investigated. This provides additional evidence on gas market integration and price correlation in two European gas markets that are distant geographically from one another.

Third, the **regression methods of least square, maximum likelihood, machine learning gradient decent** and **least square optimization** are used to compute the coefficients of a **multivariate causal regression analysis**. This study tests the short-term prediction of wholesale natural gas prices for each method used. It is found that where the linear approximation is not valid, the method suffers accordingly. However, the mathematical methods of gradient descent and least squares optimization help visualize the data sets, highlight, and accentuate the nonlinear effects of several variables on the spot gas prices. The advantage of the non-

parametric econometric model and the calibration tests used (PCA and Gamma), have contributed in accurately forecasting short term gas prices.

Last but not least, the **Information Theory** is used to assess the expressive power and efficiency of an information structure contained in the gas prices time series and check if this information is indicative and carry information that is useful to the oversight duty of regulators. Econometric and mathematical methods based on Game Theory, Records Theory, and Shannon Entropy are used to measure the following signals: level of competition, price stability, and price uncertainty respectively – for the U.S. and Europe gas markets. Second, the level of information generated by these signals is quantified using the Information Theory. The results of this innovative two-step approach show that the functioning of the European market requires the regulator's intervention via applying additional rules to enhance the competitive aspect of the market, while this does not hold for the U.S. market. Also, the value of the information contained in both markets' wholesale gas prices, although in asymmetric terms, is significant, and therefore proves to be an important instrument for the regulators.

The use of several econometric models and theories in different case studies is useful for natural gas stakeholders, as it extracts valuable information from the prices of this commodity. This information is important for traders, regulators, consumers and producers.

Table of Contents

1.Introduction 14

1.1 Gas market overview, dynamics, and development 14

1.2 Research questions and contribution of the work 22

1.2.1 World gas market, spark question – The nominal "one price for gas markets" 25

1.2.2 World gas market consequent question – The role of gas regulators 25

1.2.3 European gas market assessment I – European gas price mechanism 27

1.2.4 European gas market assessment II – Main drivers for European wholesale gas prices 28

1.3 Current state of the gas market 30

1.4 Thesis outline 34

2.Introduction review: Mathematical models used in gas markets 35

2.1 Challenge 1 – Econometric models that focus on the tail of distribution 35

2.2 Challenge 2 – Game theory econometric models 36

2.3 Challenge 3 – Econometric models that focus on the whole distribution 37

2.4 Challenge 4 – Information theory 38

3.Methodologies applied and progress beyond state of the art 41

3.1 Long term forecasting 41

3.2 Medium term forecasting 42

3.3 Medium to short term forecasting 43

4.Modeling the price dynamics of three different gas markets – Records Theory – Challenge 1 45

4.1 Nomenclature 45

4.2 Introduction 46

4.3 Material and methods 47

4.3.1 Study area and data used 47

4.3.2 Record theory models 47

4.3.3 Independent and identically distributed (*i.i.d*) case 48

4.3.4 Non - Independent and identically distributed case 48

4.3.4.1 Yang record model 49

4.3.4.2 Discrete-time random walk model (DTRW) ... 51
4.4 Results and discussion ... 52
4.4.1 Results of the three models ... 52
4.4.2 Difference between empirical and theoretical findings ... 55

5. Modeling post liberalized European gas market concentration– A Game theory perspective – Challenge 2. 57
5.1 Nomenclature ... 57
5.2 Introduction and literature review ... 57
5.2.1 Current state of the European gas wholesale market ... 57
5.2.2 Gas market characteristics in Austria and the Netherlands ... 58
5.2.3 Literature review ... 59
5.3 Material and methods ... 61
5.3.1 Parametric approach ... 63
5.3.2 Non - parametric approach ... 66
5.4 Results and discussion ... 68

6. Northwestern European wholesale natural gas prices – Comparison of several parametric and non-parametric forecasting methods – Challenge 3 ... 72
6.1 Nomenclature ... 72
6.2 Introduction ... 73
6.3 Material and methods ... 74
6.3.1 Study area and data used ... 74
6.3.2 Mathematical methods and data analyses ... 76
6.3.2.1 Least square method ... 77
6.3.2.2 Maximum likelihood methods ... 78
6.3.2.3 Gradient decent method ... 80
6.3.2.3.1 Principal component analysis ... 80
6.3.2.3.2 Gamma test ... 81
6.3.2.4 Least square optimization methods ... 88
6.4 Results and discussion ... 89
6.4.1 Results of OLS method ... 89
6.4.2 Results of the maximum likelihood method ... 90
6.4.3 Results of the Gradient decent method ... 92

6.4.4 Results of the least square optimization method 93
6.4.5 Comparison analyses of all models 94

7.Modeling the price of post liberalized gas markets – Information theory – Challenge 4 97
7.1 Nomenclature 97
7.2 Introduction 97
7.3 Material and methods 101
7.3.1 Data 102
7.3.2 Signal 1 – Level of competition 103
7.3.3 Signal 2 – Market stability 104
7.3.4 Signal 3 – Volatility and uncertainty in prices 104
7.3.5 Information theory 106
7.4 Results and discussion 108
7.4.1 Results of the non - parametric Cournot test 108
7.4.2 Results of the Records theory 109
7.4.3 Results of the Shanon entropy 111
7.4.4 Synthesis of signal results 111
7.4.5 Results of the Information theory 113

8.Synthesis of results and conclusions 118
8.1 Findings, discussions and conclusions referring to the research questions 118
8.2 Strengths and weaknesses of each challenge/ model 122
8.3 Econometric models perspective and outlook on future of gas markets 129

References 133
Appendix to Chapter 4 141
Maximum Likelihood estimation method 141
Goodness-of-fit for a Yang model 142

Chapter 1

1. Introduction

1.1 Gas market overview, dynamics, and development

The natural gas remaining resources[1] are sufficient to easily meet the global demand growth projections to 2040 and beyond [1,2]. Proven reserves stood at around $200\ Trillion\ Cubic\ Meters\ (tcm)$ at the end of 2019, approximately equal to 60 years of production at the current output rates. Russia is the leading country with $48.7\ tcm$, followed closely by Iran and Qatar.

The global gas reserves expansion is mainly driven by the profitability from the unconventional gas that is yet to be produced in North America and Asia, as well as conventional gas in Russia and the Middle East.

The United States, Russia and Iran are the three largest gas producers today; this ranking is likely to remain unchanged until 2040. Data [2,3] show that the United States produced $870\ Billion\ Cubic\ Meters\ (bcm)$ in 2019, followed by Russia and Iran, both producing, $650\ and\ 230\ Bcm$, respectively. With $848\ bcm$, the United States is also the world's largest gas consumer, followed by the European Union and Russia, both standing at $460\ and\ 505\ bcm$ respectively.

Gas demand is expected to expand almost everywhere in the coming 25 years – according to [2] the European Union and Japan are the main exceptions – however the growth will clearly be concentrated in developing countries. As with oil, the gas market is characterized with geographical inequalities between the volumes of gas produced and consumed. Gas trades should fill this deficit.

The cross-national gas trade has been growing for the past years because of the recent Liquefied Natural Gas (LNG) exports development and the ongoing construction of new long distance pipelines. Today, 30% of produced gas is traded across national boundaries, two-third by pipeline and the remaining third by LNG. Top gas exporters are Russia, Norway, Qatar, Australia, and, the United States [3].

The main parts of the world that are witnessing substantial gas deficit and relying on gas imports are the European Union ($329\ bcm$), followed by the two developed countries of Japan (165 bcm) and Korea (165 bcm), and finally China ($73\ bcm$). The major trade flows are shown in Figure 1.

[1] A resource is the amount of a geologic commodity that exists in both discovered and undiscovered deposits—by definition, then, a "best guess." Reserves are that subgroup of a resource that have been discovered, have a known size, and can be extracted at a profit

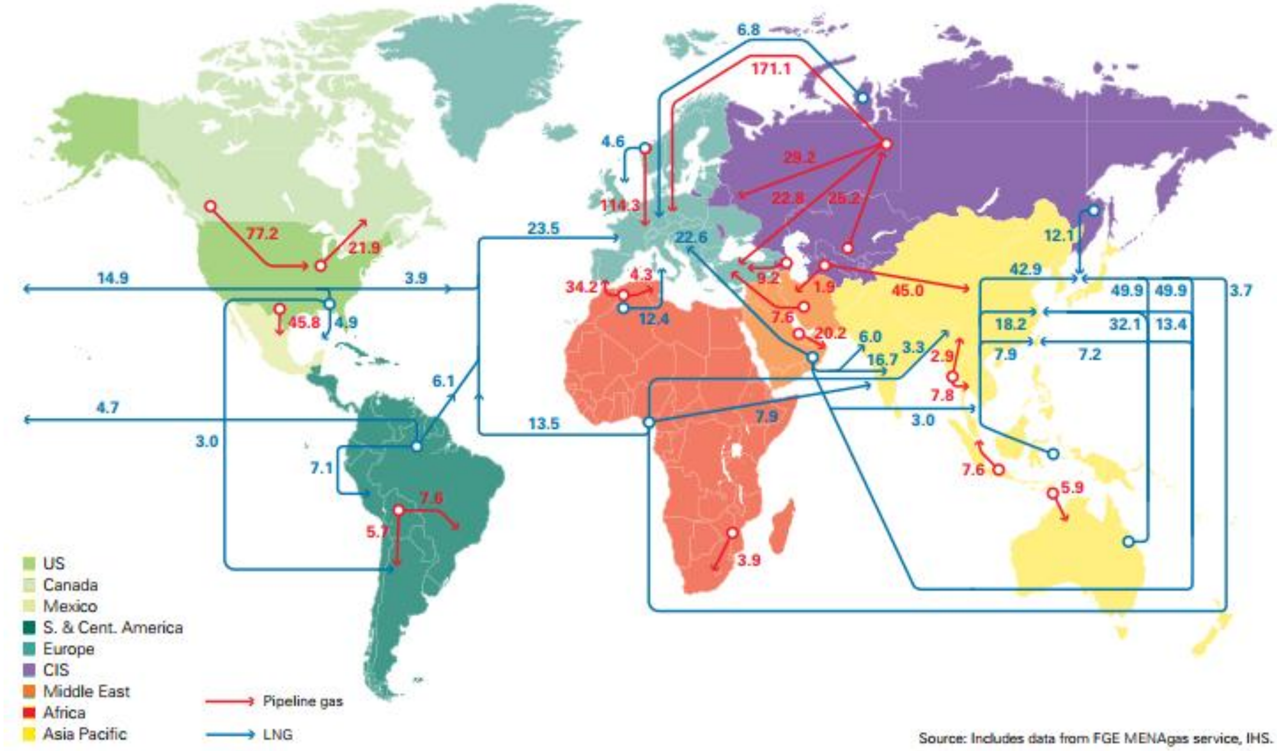

Figure 1. Major trade movements 2018 - Trade flows worldwide, billion cubic meters

In terms of LNG exporting countries, Qatar has emerged in recent years as the leading country worldwide. In 2016, Qatar's exports accounted to $1/3$ of global LNG exports, followed by Australia (11%) and Malaysia (10%).

In terms of total LNG demand worldwide, there are seven LNG importing countries that cover almost 80% of total volumes of the global LNG market. In fact, the top-3 LNG importing countries are entirely Asian countries, namely: Japan (34%), South-Korea (13.2%) and China (7.9%), [4].

The first part of the section introduced statistics regarding the gas market resources, supply and demand; the remaining part of this section will explain the Gas pricing system, followed by a short discussion on the new gas market structure.

Gas pricing system does not follow one specific function and is consequently not valued in the same way worldwide. The gas industry originally developed as a series of isolated regional markets, often with their own pricing systems:

- Oil price indexation: Gas prices are set by a formula that is indexed to oil prices under long-term contracts
- Bilateral monopoly: The dominant pricing mechanism in deals involving the former Soviet Union, central and Eastern Europe, and china

- The other categories apply to internal markets and most commonly are used to subsidize domestic customers.

In recent years, however, industrialized countries are undergoing profound structural changes brought about by governmental policies aimed at liberalizing the existing gas markets; these policies will introduce a new inter-commodity competition based on third party access to gas supply infrastructure, privatized public gas utilities, and enforced new contract mechanisms such as spot markets, short-term and non-dedicated contracts.

Consequently, a cross border gas trade between neighboring markets trade began to develop, for example, the recent development of European and Russian gas pipelines, and for the emerging role of LNG in recent years. Thus, international gas markets have experienced new pricing mechanisms that consist of:

- Full gas to gas competition, gas priced in open free market trade
- The addition of price revision clauses to classical long term oil price indexation contracts giving both parties the right to request price revisions to reflect changes in the gas market price.

The global natural gas market is comprised of regional markets that are often grouped based on the regions of natural gas trade. In recent years, roughly 70% of global natural gas trade has been transported to market destinations within the country of production, while the remainder has crossed international borders: through either long-distance pipelines or liquefied natural gas [5].

Three regional gas markets exist around the world; each one is clustered by a regional spot market (Figure 2):

- The US market – cleared by Henry Hub spot prices
- The European market – cleared by spot prices at European hubs
- The Asian market – cleared by spot LNG prices

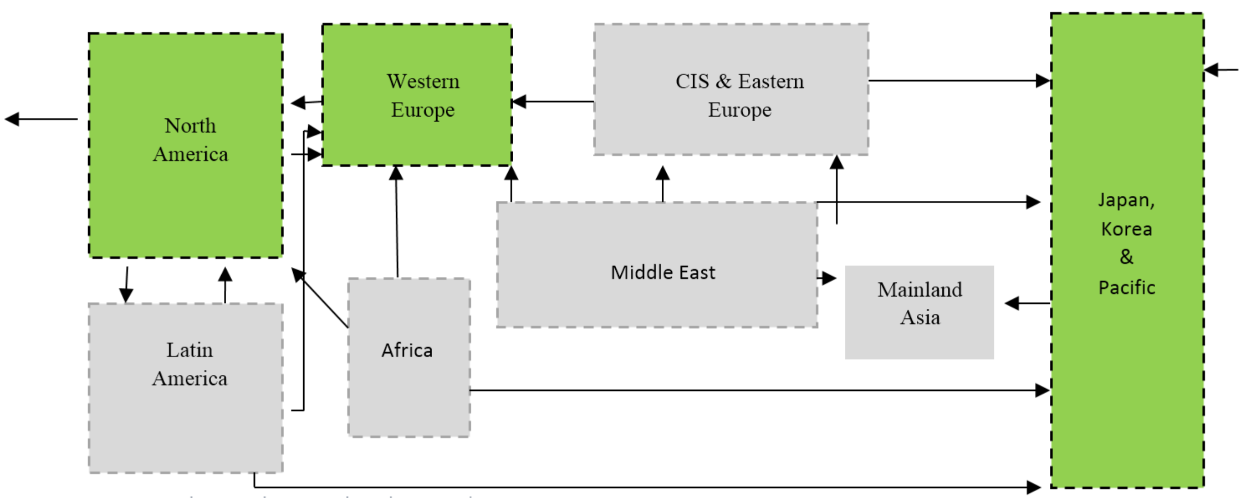

Figure 2. Main regional Natural gas markets (in Green)

The most liberalized system is in North America, where gas pricing is highly competitive and based on supply/demand balances (gas to gas competition). Asian and European gas prices, however, are still influenced by oil-indexation.

In the European and the American regional markets, natural gas is mostly purchased through pipelines because of large domestic resources and strong transmission grids. The lack of such infrastructures in North East Asia prevents the import of natural gas through pipelines. Therefore, natural gas could only be imported in the form of LNG.

In Northeast Asia, for example, almost all LNG contract volumes are indexed to crude benchmarks (e.g. Japanese Crude Cocktail, Brent) because of their dependence on external imports [6]. However, as long-term contracts expire, the oil-indexation system is eroding.

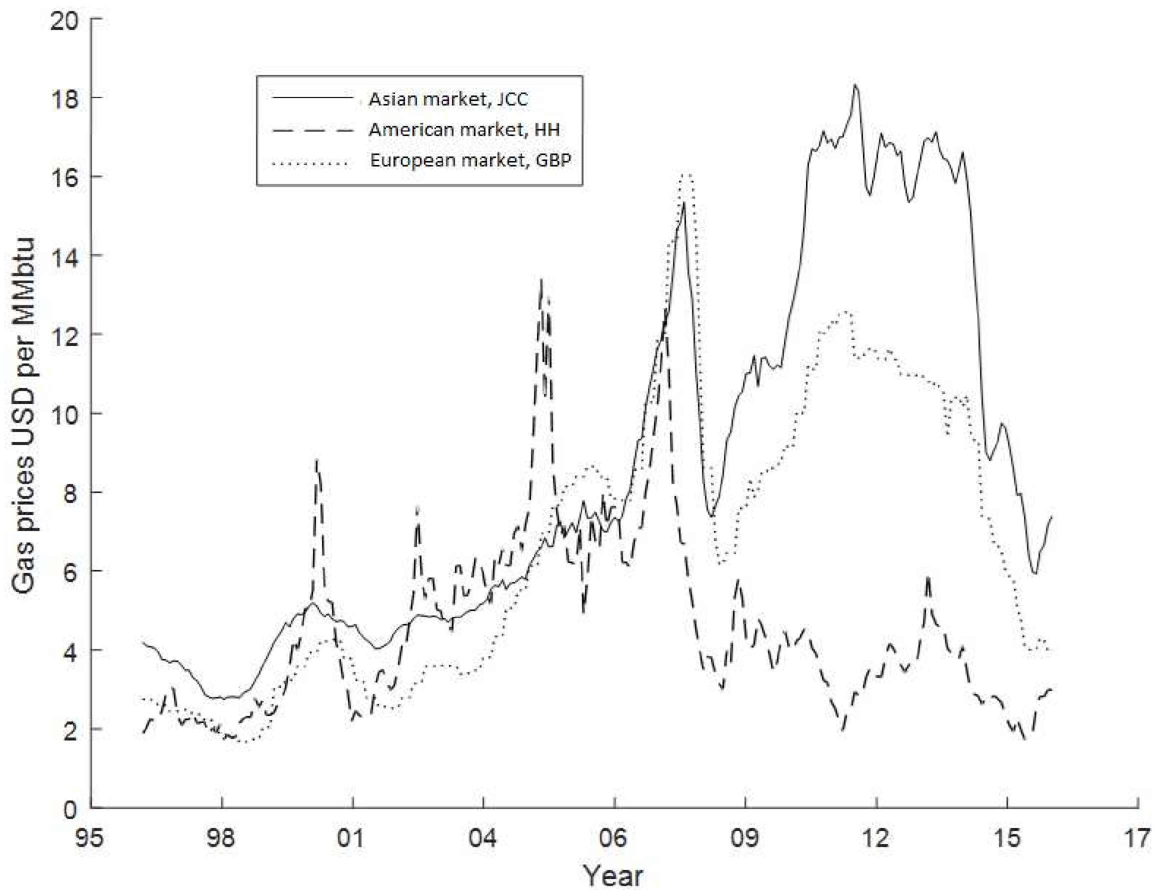

Figure 3. Monthly Gas prices in three different Natural gas markets

Different gas hubs characterize each gas market. A hub is a platform where the title (ownership) of gas molecules is exchanged between a number of buyers and sellers in both spot and futures trades [7]. The major parties involved in the development of a natural gas hub include market participants, transmission system operators, hub operators, brokers, and energy companies, among others.

According to [8], the creation of a functional gas hub requires market liberalization and transition of gas pricing mechanism. Once created, the hub will be developed by a political will and regulations that are essentially needed to further safeguard the competitive environment.

In order to show the readers the gas price evolution in the three different gas markets, the most important hub/ price reference in each region will be selected. Figure 3 shows the time series of natural gas prices of the three different regional markets. The data set consists of monthly values recorded from March 1997 to November 2016. The time frame will be divided into five main periods, and a brief explanation of the fundamental supply and demand factors that led to the price behavior will be given. It is worth noting that this long period has witnessed lots of developments and changes, as several new gas hubs have emerged.

- The German Border price (GBP) is a useful reference price in Europe, as most of the producer/wholesaler gas imports (mainly used in the old point-to-point system) used the German border as the transfer of ownership point. The German Border Price (GBP) is an average of the oil-indexed contracts, and is comprised of a big percentage of Russian gas supplies and spot supplies. The latter is increasingly available at the Dutch-German border and Norwegian pipeline terminals.
- Spot and future natural gas prices set at Henry Hub are the primary price set for the North American natural gas market [9]. The Henry Hub, lends its name to the pricing point for natural gas futures contracts traded on the New York Mercantile Exchange (NYMEX) and the OTC swaps traded on Intercontinental Exchange (ICE).
- Japanese Crude Cocktail (JCC) is the average price of customs-cleared crude oil imports into Japan (formerly the average of the top twenty crude oils by volume) as reported in customs statistics. JCC is currently used as the index of gas prices in the Asia Pacific, where Japan's shares alone constitute 37.2 percent of the total imports in 2013 [10]. The Asian LNG value of imports, in particular, depends on the demand for gas and the long-term contract price (a value that is based on the JCC oil-linked pricing formation) [11].

<u>1997-2006</u>

Gas prices in Asia and Europe moved most closely with crude oil prices, as these were formally indexed to crude/ fuel products. Wholesale gas prices in the United States were more volatile, but they, similarly, tended to co-move broadly with oil prices on accounts of switching between energy types (natural gas and fuel oil).

During this period, the U.S. natural gas market was the world's largest, with 75 percent of the supply produced domestically and the balance between storage withdrawals and imports from Canada. Tight supply-demand balance has characterized U.S natural gas markets, and has led to natural gas spikes in several periods during 2000-2008. As shown in the graph, this period witnessed an upward thrust in US gas prices.

The upward of price trend was also seen in Asia and Europe. However, no shocks were observed in the U.S. The price differences between JCC and GBP are mainly due to two reasons: first, Liquefied Natural Gas import prices in the Pacific Basin are more expensive by roughly $one\ USD\ per\ MMBtu$ than those of the Atlantic Basin[2]. Second, the gas market in continental Europe does not heavily rely on gas imports via LNG.

2006-2008

From 2007 into the first half of 2008, natural gas prices maintained a rising behavior in all markets. This is mainly due to tight supplies and unprecedented growth in oil demand. As mentioned previously, oil-indexation in Europe was dominant in international trade; however, since 2006, the use of gas-on-gas pricing steadily increased to reach a significant percentage. In Asia, oil-indexation was also dominant, but unlike in Europe, has not shown steady decline.

2008-2010

A dramatic drop in prices occurred in 2008, when the depth and spread of the global recession became apparent. Gas prices in key liberalized markets in the US fell from $USD\ 13-14\ per\ MMBtu$ in mid-2008, to a minimum of $USD\ 4\ per\ MMBtu$ in April 2009. In the second half of 2009, natural gas prices and crude oil prices in the US have stopped co-moving with other markets. This could be interpreted as a result of the growing shale gas production. As an oversupply of natural gas in the Atlantic basin was created, the gas-on-gas prices dropped regardless of the global economic recession.

Prices in markets, such as Japan and Europe, linked to oil dropped from their peaks as these prices have lags of 3 to 6 months [12]. Greater regulatory activity was critical to open up cross-border activity. Improvements in hub trading were important because they refined price discovery, and as a result ensured competitively priced gas and enhanced energy security.

2010-2014

Oil linked gas prices in Asia increased from 2010 and onwards to above $15\ USD\ per\ MMBtu$, bearing the legacy of the above 100 USD oil prices. In Japan for example, wholesale gas prices continue to co-move with oil prices, although the indexation with oil prices may come under pressure as spot markets in Asia start to develop.

The importance of natural gas Spot markets has grown steadily in Europe. This growth, combined with high oil prices and low natural gas demand after the economic crisis, has led to a renegotiation of many indexed contracts, linking new contracts to spot markets (so-called "gas-to-gas pricing"). As shown in Figure 3, gas prices in Europe have increasingly decoupled from oil prices and did not increase as strongly as oil prices between 2009 and 2011.

[2] This premium (the "Asian premium") is due to long-haul shipping of gas, high charges applied to the use of LNG terminals and lastly the absence of competition from piped gas.

2014-onwards

Spot LNG prices decreased significantly in 2014 and in early 2015 both in Asia and Europe. The reasons for the decrease are a weak demand in Asia, an increase in global supplies, and a drop in oil prices. In recent years, the average German border price followed predominately the gas to gas pricing mechanism [13].

Estimated border prices showed a clear declining trend over 2015 and early 2016. Driven by the oil price drop observed in the second half of 2014, oil-indexed prices fell faster than hub-based prices. Consequently, a significant price convergence in the third quarter of 2015 between the Asian and European gas prices was observed.

As such, natural gas, which is traded on the wholesale market, exhibits particularly large increases in price volatility. The rise of competition and deregulation leads to relatively free energy markets, which are characterized by high price shifts. Therefore, the market is “vulnerable” to price spikes/drops. Records of historical statistical data sets of variables recorded at successive periodic intervals constitute the first step in forecasting, which is the subject of this thesis.

The introduction started with an overview of “gas supply and demand”, a section that focuses on the dynamics of the gas prices in three different regions, and will be concluded with a section that briefly focuses on the European region (the major region of focus in this thesis)

The global spatial distribution of proved gas reserves in Europe is shown in Figure 4, [14]. It impressively demonstrates why Russia is on the top of the list of the most important gas exporting countries to Europe. Africa is an important source of pipeline-based gas imports to Europe as well. Additionally, there are 16 LNG regasification facilities located in Western Europe and 40 to 50 others under consideration and/or construction as of 2015. Imported gas is transported across the European continent by a vast pipeline network.

So far, pipeline-based gas supply has been the unconfined technology to transport gas from the source to the sink, at least on the border to Europe. Therefore, geography has played a key role in European gas supply, which can be described as follows:

- North-East: Russia covers the highest share of pipeline-based European gas supply.
- North West: Covered by the two main EU gas producers of UK and the Netherlands, in addition to the main gas imports from the EU trade partner Norway.
- East (Middle East, Central Asia): Countries like Iran, Iraq, and Turkmenistan are potential candidates for imports. Due to different reasons, however, none of them is a significant gas importer to Europe (Iran: trade embargo, political reasons; Iraq: war, political reasons; Turkmenistan: proven gas reserves have been made public a couple of years ago).
- South-East (Gulf Region): Europe imports to a certain extent LNG from the Gulf region, notably Qatar.
- South (North Africa): Algeria is the main gas importer from Africa, followed by Libya. Europe also imports LNG from Nigeria.

- Remaining directions outside Europe (Overseas): LNG imports from countries like Trinidad, Peru and other countries overseas.

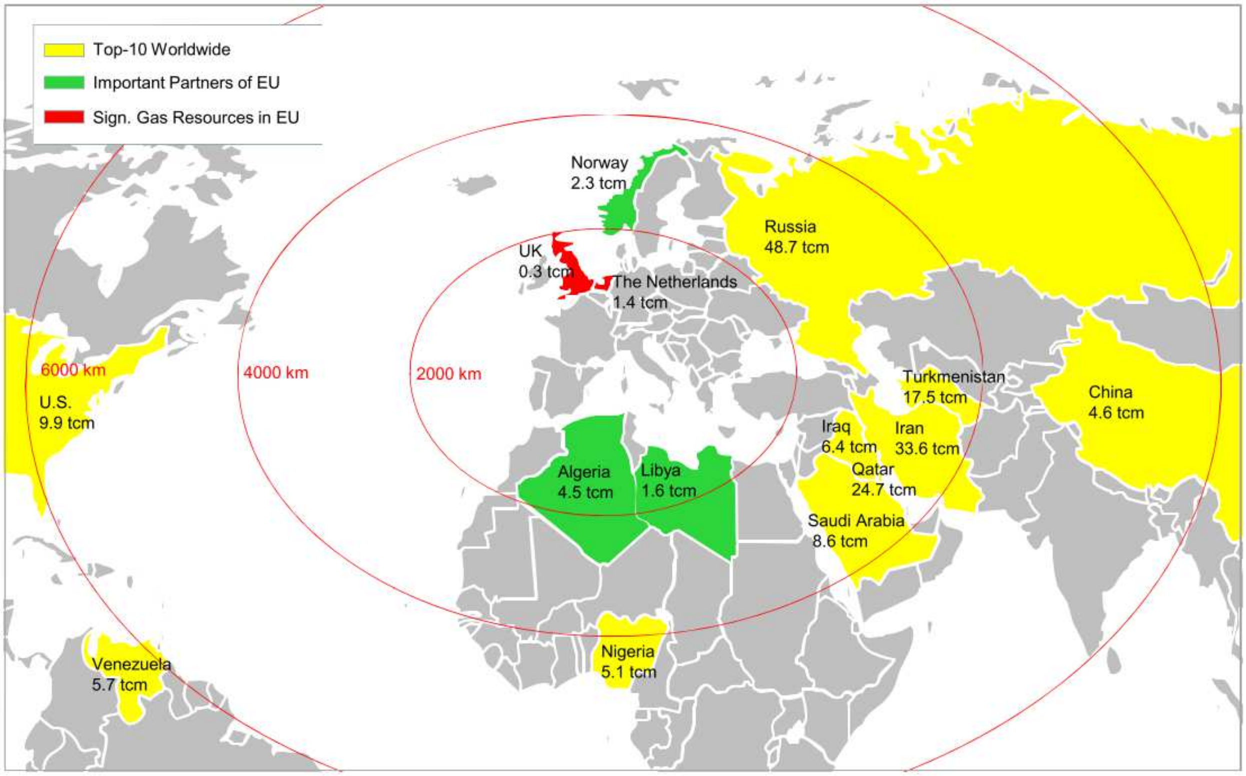

Figure 4. Global spatial distribution of proved gas reserves relative to Europe

This gas supply covers annual European gas demand that was – on average – around $545\ bcm/year$ from $2005\ to\ 2010$. Afterwards, annual European gas demand continuously decreased mainly as a result of the financial crises, mild winter seasons, lower gas demand in electricity generation due to increased share of renewable technologies. The demand in 2019 was $482\ bcm/\ year$.

1.2 Research questions and contribution of the work

There are many reasons why a better understanding of the stochastic process driving the prices of natural gas would be useful. This understanding would be helpful on the microeconomic level, including but not limited to: Understanding the main factors that affect the supply and demand of the commodity, providing an efficient tool for better forecasting the gas prices, aiding with the decision, and timing of new gas infrastructure, etc.

In this thesis, a wide variety of important econometric methods will be used and linked to a common "use case" in the subject of commodity pricing, such as natural gas. Classic econometric methods, such as the parametric and non-parametric forecasting methods, focus on the whole distribution and are used in a multivariate causal regression analysis to forecast gas prices. Besides, the Records theory that focuses on the study of extremes on one hand, and the complementarity Games Theory on the other were also used to study several aspects of the gas markets (not limited to the level of competition, long term price stability, etc.). Finally, and not drifting apart from probability and statistics, Information Theory will be used to extract indication about the nature of markets, that can be used by gas regulators in their quest for market oversight.

From a mathematical point of view, I am applying a battery of techniques (econometric models) that has not been used before in the natural gas markets scientific literature. The choice of the relevant econometric model is not random and is derived from three main pillars: Research question, data and theory.

The research question is the spark and starting point of choosing the appropriate model. This will be later explained in sections 1.2.1 to 1.2.4. For example: if the intention is to forecast gas prices time series, I have to look at parametric and non-parametric regression models, it starts with normal least square methods and does not end up in non-parametric machine learning methods. On the other hand, if the aim is to compute probabilities of future price spike or drops, I have to rely on other models, such as the use of Theory of extremes, etc.

The data is essential, and is a main parameter that needs to be considered when choosing appropriate econometric models. For example: a data that consist of many outliers cannot be modeled as stationary, instead random walk models can alternatively be used. Other modelers might decide to remove outliers completely. This comes at a cost, simply because outliers can contain more information content than other data points. On the other hand, selecting a model, completely disregarding the data is not appropriate, as it will certainly lead to wrong result misinterpretation.

The theory consistency is another important pillar, and needs to be fully understood before selecting the appropriate model. For each econometric model, there exists a set of mathematical/ statistical assumptions. For example, a non-linear complementarity problem containing a set of constraints, could have a convex set with a convex objective function, which will eventually lead to find a global minimum for the problem. However, this is not the case in non-convex problems. Therefore, the modeler should really know how the model is built and the theory behind it.

Another part of the **theory** is the model validation. If the model results cannot be validated, this means that the model was not successful. An example to illustrate this point can be given for parametric model of least square and maximum likelihood that rely on a set of assumptions, such as the homoscedasticity, non-autocorrelation and normality of residuals. If the results do not respect and validate such assumptions, therefore the models cannot be validated.

From an economic point of view, this thesis is built open addressing several challenges and is divided into two major parts, each of which contains a set of open-ended questions. The aim is to find a suitable econometric model, transform the challenge into a specific research question, and try to answer it empirically and theoretically.

The first part addresses the global gas market, worldwide gas prices convergence, and whether the various gas markets are closely correlated[3] and react in a similar manner to changes in supply and demand fundamentals from one side; and the role of gas regulators in such markets from another.

While the second part focuses on and the assessment of the European gas hubs. The main intention is to test the market concentration and check whether the liberalization process has led to fair competition between the different suppliers from one side, and whether the liberalization has contributed in creating one European gas market, where price difference is only limited to transmission fees. Additionally, a closer look on the main drivers of wholesale prices, price unpredictability and a forecast of short term gas spot prices will be addressed.

An illustration of the challenges is presented in Figure 5 below.

[3] Taking into account the major liberalization reforms that are happening in various Gas regions.

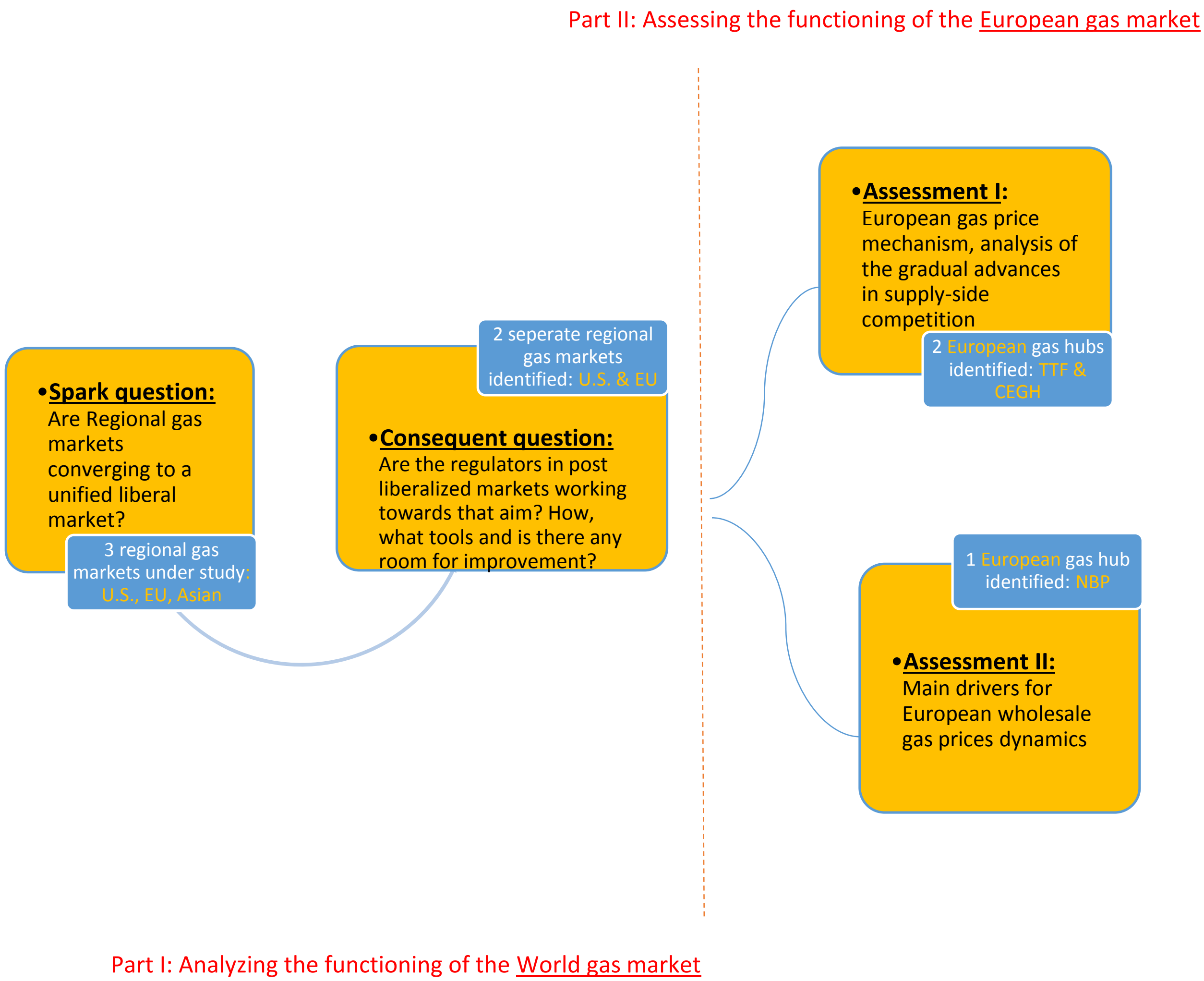

Figure 5. Research questions and challenges

The figure is split into two parts, the left half displays the research questions related to the world gas market, the price convergence to a nominal "one price" and the role of the regulators, while in the right half the focus is shifted to the European gas market. Two assessments are conducted for the latter market, the gradual advances in the supply-side competition, otherwise known as the level of competition between gas suppliers/ wholesalers, and the drivers of gas prices for European gas hub. The reasoning behind this partition and how one thing led to another is briefly described in the following sections:

1.2.1 World gas market, spark question – The nominal "one price for gas markets"

The spark of my research is motivated by two main factors. First, gas trade has been limited geographically within three main regional gas markets: Europe, North America, and Asia Pacific. The lack of transmission infrastructure limits the trade. Unlike the oil market, the gas market is not globally connected; this implies that each regional gas market is characterized by specific supply and demand conditions.

Second, there are times where wholesale gas prices exhibited in the three different regions are remarkably different from each other's, and the price difference can reach significant values, hindering the development of such a commodity in some places while flourishing in other parts.

The increase in infrastructure investment and additional liberalization could pave the way for additional integration of world natural gas markets, and possibly lead to more competition between suppliers, and bring more stability to the gas prices worldwide. Price stability and strong market interconnections might be two of the main pillars that will close the gap between the gas prices found in different gas regions, on the hopes that it converge towards international gas prices that are closely correlated, quite similar to the oil market. Therefore the question to be asked is:

- **Will the global gas prices head to a nominal "one price" and consequently will all buyers have to pay that price?**

The Records theory will be used to study the effect of extreme gas prices and to test the stability of the three different regional gas markets. Records theory studies observations that are higher than all previous ones, which is equivalent to say the maximum / minimum observation up to present time, and those records, are concentrated in the tail of a given distribution, i.e. study of extremes.

In this study, several models are developed to test and analyze the stability of three main regional gas markets (U.S, Europe and Asia). Such models, compared to previous methods used in literature, have a major advantage in that the results are distribution-free. Consequently, the applicant will not be concerned about identifying the distribution type, and the models complexity is reduced.

The spark question will be referred as "challenge 1" in the remainder of this document, and the corresponding journal article is accepted and published in the *Energy strategy reviews journal*, Elsevier [15].

1.2.2 World gas market consequent question – The role of gas regulators

Having a liberalized market, does not ensure that the level of competition between gas suppliers is at its best, nor does it imply that the gas prices in such markets are prone to drastic and sudden changes in the short and long term. This kind of information is essential to all participants in the gas value chain. Among all gas stakeholders, this information is important for gas regulators. The two defined states in this study are either for the regulators to take action or keep the business as usual (BAU). The former case is applied whenever there is a possible indication of market abuse by some gas suppliers. In such a case, the regulator needs to intervene, to adjust the legal framework along with the supply and demand fundamentals.

It is clear to the author that out of three regional gas markets addressed in challenge 1, the Asian market has not witnessed much liberalization in the recent years, and is still facing major problems on several levels: legislative, supply/ demand level, infrastructure, etc. In a nutshell, the Asian gas market, represented by its most liquid hub, Japan Crude Cocktail, JCC, is still lagging behind in terms of liberalization when compared with other regional gas hubs. Therefore the Asian market is omitted in this part and only the U.S. and the European gas markets will be addressed.

In order to understand the degree of liberalization and assess the performance of two of the most liberalized gas markets, the following research questions should be answered:

- **How can the use of statistical and probabilistic theory be used to extract information and signals about gas market performance that could be used by gas regulators?**
- **Can this signal be trusted?**
- **Is there a need for further adjustment in the legal framework of the gas markets in the U.S. and Europe?**

The level of competition changes from one market to another, and if measured correctly defines the concentration of competing firms in the market. Few and a limited number of firms imply a highly concentrated market, and that based on their strategies can dictate prices, otherwise known as price-setters. Besides the fewer the number of firms the easier it is to abuse conduct and act collusively. Such firms adjust their strategies in conjunction with an agreed-upon understanding with the competing firms at the expense of the welfare of gas consumers and possibly smaller firms. A typical example of such a market and behavior is the presence of cartels in commodity markets.

The level of volatility indicates how fast and sudden gas prices change in the short term. The higher the volatility the harder it is to predict the future behavior of the changes, thus making the market uncertain.

Price stability hint at the behavior of gas prices in the medium and long term on the other hand. Commodity prices tend to have abrupt and rapid price shocks, and this is witnessed when prices suddenly increase or decrease due to sudden changes in supply and demand characteristics. The longer it takes for a commodity price to witness a shock the more stable the market is.

The performance of the market is the measure of the power and efficiency of the information contained in the gas prices and the signals. The more efficient, the more reliable, and reflective the prices are in such a market.

The first objective of this section is to study whether the wholesale gas prices of two of the most liberalized gas markets carry valuable information which can serve as signals for the relevant gas regulators. The value of these signals will be quantified by using several econometric methods and mathematical theories. This analysis will guide and assist the decision-making process of regulators regarding the need for an intervention to stabilize the gas markets and improve the functioning of their internal markets. The second objective is to quantify and measure the accuracy and efficiency of the hidden information structure generated by these signals.

The consequent question will be referred as “challenge 4” in the remainder of this document, and the corresponding journal article is accepted and published in the *energies journal*, MDPI [16].

1.2.3 European gas market assessment I – European gas price mechanism

It is obvious from the results of sections 1.2.1 and 1.2.2, that the U.S. market is the most liquid and liberalized in the gas world. This is due to several facts that are described in the next chapters. On the other hand and as mentioned in section 1.2.2, the Asian market is lagging behind in terms of liberalization. This motivated the author to shift its focus on the European gas market in the second part of this thesis.

Compared to the previously used point-to-point system adopted in previous years in Europe, the new entry-exit model represents a general improvement providing more flexibility for network users on the basis of non-discriminatory access, fostering competition, and creating an EU internal gas market for natural gas. Several barriers still persist; nonetheless, the answers to the research questions below might validate the new entry-exit model improvements and highlight what needs to be addressed to have a unified European gas market in the near future:

- **Long-term contracts are a cornerstone of the energy security in Europe, and oil-indexation is a hedge against price manipulation by dominant suppliers, but do these claims still hold true?**
- **What is the future of long-term contracts?**
- **Is gas indexation a solution or a recipe for disaster?**
- **Will the correlation between oil and gas cease to exist in such markets?**
- **Is there an integrated gas market in Europe?**

In recent years, a shift towards gas-on-gas competition can be observed in Europe [17]. The pressure to liberalize European gas markets has partially broken the link of oil and gas prices with gas increasingly traded separately on the basis of supply and demand [18].

According to Table 1, the gas-on-gas competition index contracts in Central and Eastern Europe has increased from zero in 2005 to over 56% of the total contracts between suppliers and consumers, with changes rising from 2011 onwards. The zero simply means that the gas prices were totally indexed on oil prices. The change is mainly due to the liberalization process and the establishment of the future and spot market covering the three most important hubs in continental Europe, GASPOOL, NET-CONNECT Germany and the Dutch Title Transfer Facility (TTF) by the European Energy Exchange[4]. Nowadays, market pricing stretches into the rest of Europe through interconnected pipelines with major hubs [19]. However, the indexation level in the North-Western part has reached higher levels [20].

Region	2005	2015
Northwest Europe	27	92
Central and Eastern Europe	0	56

Table 1. Indexation level of European gas contracts

[4] 2011, EEX introduced 24/7 trading on the Natural Gas Spot Market. Short term gas quantities can be traded on the exchange for delivery in the market areas GASPOOL, NET-CONNECT Germany (NCG) and the Dutch Title Transfer Facility (TTF) area.

By 2013, the majority of consumers in the largest EU gas markets had an increasingly credible choice of suppliers. Moreover, competition has become fierce particularly for large customers in many countries [21]. Even though, the market has changed over the years, but does the change occur in all of the European gas hubs?

To answer the above research questions, parametric and non-parametric Game Theory methods will be used to study the effect of market concentration on gas prices. Austria and the Netherlands will be the focus since each represents a different evolutionary stage in the process of wholesale natural gas markets liberalization.

The parametric method takes into account the classical Nash-Cournot equilibrium test, with assumptions on cost and demand functions. However, the non-parametric method does not make any prior parametric assumptions, a factor that allows greater freedom in modeling. The advantage of using Game Theory on the supply and demand side of the European gas markets lies in the quest to identify whether the liberalization process has achieved the goals of reducing price manipulation in Europe as a whole from one side, and increasing market integration from another.

This assessment of the European gas market will be referred as "challenge 2" in the remainder of this document. The final challenge will address additional research questions related to main drivers of the wholesale price variations and short-term price forecast.

1.2.4 European gas market assessment II – Main drivers for European wholesale gas prices

As mentioned previously, the gas indexation level in North Western Europe has reached higher levels in recent years, therefore gas prices are less affected by oil prices, and more affected by supply and demand factors. In order to understand the main drivers behind the gas prices behavior in a comprehensive and analytical way, statistical econometric models will be used in quest to answer the following research questions:

- **What drives natural gas prices in European hubs?**
- **What inputs are relevant to make the prediction and which ones are irrelevant?**
- **How many data points are required to make a prediction with a best possible accuracy?**
- **Which optimization technique is more adequate to use?**

Based on the data of natural gas' weekly prices from September 2007 to December 2014 of the German gas hub NCG (Net Connect Germany), and variables, such as substitute fuel prices (coal and oil), weather data, exchange rates and storage utilization rates, the methods of least square, maximum likelihood, machine learning gradient decent and least square optimization, are used to compute the coefficients of a multivariate causal regression analysis. The study also tests the short-term prediction of wholesale natural gas prices for each method used.

Classical regression models are used to compute the coefficients of the parametric regression methods, namely multi-linear regression and Vector Autoregressive Analysis. The best choice for the application of the non-parametric methods is the machine learning feed-forward multi-layer perceptron neural network. A frequent problem of machine learning, caused by a miscalculation of the number of observation and variables, is over fitting. In order to overcome this level of complexity, several statistical techniques were used.

Firstly, the main linear technique for dimensionality reduction, Principal Component Analysis (PCA), performs a linear mapping of the data. Secondly, the Gamma test, which is a mathematically proven smooth non-linear modeling tool, is used to choose the best input combination before calibrating and testing models, therefore reducing the inputs selection uncertainty.

The above assessment will be referred as "challenge 3" in the remainder of this document. The results of such a study will enhance the understanding of what causes natural gas price volatility in liberalized gas hubs in the first place, and attempt to accurately forecast short term gas prices in a second place. This will be done by a comparing the performance of the different econometric models described above.

The corresponding research is accepted and published (in press) in the *International Journal of Global Energy Issues,* Inderscience [22].

1.3 Current state of the gas market

The four challenges listed in this document are sufficient to stimulate, investigate, and assess the progress of the gas market. This in turn leads to an increasing appreciation and importance of the relevant modeling techniques. A review of the statistical and mathematical modeling progression applied in the field will follow in chapter 2. This section will list the findings and results of previous studies that tried to answer some of the research questions cited previously and highlight the gaps.

Challenge 1 – The nominal "one price" for gas markets

As previously stated in section 1.1, natural gas markets have developed drastically in recent years. Ongoing liberalization process worldwide, reforms of energy and utility companies, technology progress, and recent trends in LNG market have all contributed and pushed closer the three distinct and segmented gas regions. The most notable changes are apparent in new contract models and pricing mechanisms [23].

Researchers, such as [20,21,24–28], agree on the fact that out of these three markets, the most liberalized and stable market that reflects gas to gas competition is the American gas region, followed by both the European and the Asian markets that are less integrated and more prone to high prices and unpredictability. Developments in regional demand patterns, new offshore discoveries in virgin areas, additional connectivity either by pipelines or by LNG might lead to more or less stable markets, where price differentials between the different regions is low.

In the quest to validate the opinion of the cited authors/ studies, and to answer challenge one, the Records Theory will be used and further discussed in chapters 2 and 4.

Challenge 2 – European gas price mechanism

A characteristic feature and offspring of the liberalization process that has been implemented in Europe is the emergence of trading hubs. Unfortunately, there are some areas where the competition is minimal, thus dominated by a few traders and customers on both the supply level and consumption level.

The dynamics of European gas markets and the effect of the liberalization process have received increasing attention from research institutes and companies. In fact the Oxford Institute for Energy Studies (OIES) has developed a whole gas program to study the development of European traded gas hubs [20,29–31]. Other studies, such as the one published by DNV KEMA [32], have been carried out by the European Commission, who aims to analyze the implementation of entry-exit systems for gas in the Member States of the European Union, more specifically, the capacity products and pricing.

Two additional studies have been conducted by the International Gas Union [25] and the Agency for the Cooperation of Energy Regulators [33] on the gas price formation and the functioning of the gas wholesale markets in several members states in the European Union.

The analysis led to the following conclusions:

- The traded volumes, compared with the size of the market, are a main indicator of hub liquidity, and accordingly, the most liquid hubs are located in North Western Europe. Unfortunately, this is not the case for other hubs located more to the East, where oil price indexation is still dominant in gas contracts.
- In fact, companies, such as Statoil, Gasterra, and UK producers, shifted their gas contracts to hub indexation. Gazprom, Sonatrach, other key producers and several LNG exporting companies prefer long-term bilateral contracting with a higher presence of oil-price indexation [33]. The mentioned studies show that the effects of the liberation process have a positive impact in Europe. More specifically, the north-western part of Europe reflects gas market supply and demand fundamentals. However, this is not yet the case in the central part of Europe, namely the Austrian CEGH gas hub [25,34,35].
- In general, the behavior of the market participants in the equilibrium model can range from perfect competition to monopolistic, oligopolistic, and Cournot players with conjectured supply functions relative to their rivals[5]. Most of the researchers [36–41] agree that the Cournot competition is the most elastic representation of today's gas market, especially in the European gas market.

In the quest to validate the opinion of the cited authors/ studies and in order to answer challenge two, the study will use the Game Theory methodology and will answer whether the liberalization of the Gas industry has led to less concentrated markets in the various European gas hubs. Chapter 2 and 5 will expand on the mathematical methodology.

Challenge 3 – Main drivers for European wholesale gas prices dynamics

The majority of international gas trade outside North America is still conducted on the basis of 10 to 30 year oil-indexed contracts with complex price clauses. In 2010, the European oil-indexed prices played a dominant role, while the hub prices played a balancing/arbitrary and subordinate role.

The pressure to liberalize European gas markets has partially broken the link of oil and gas prices with gas increasingly traded separately on the basis of demand and supply [18]. In recent years, a shift towards gas-on-gas competition can be observed in Europe. The latest data for the total gas priced in the whole of Europe for the year 2014 refer to 61% to be hub-based and 32% to be traded under long-term oil-indexed contracts [17]. In the UK and the Netherlands, the spot wholesale gas markets play a central role in the overall transactions of gas, as they offer a second source of gas provision as an alternative to the traditional long-term contracts [42]. Germany and Italy, on the other hand, although recording the largest gas consumptions in Europe, are trading in their hubs a smaller fraction of the overall consumed gas that, for the most part, is bought through long term contracts [43].

According to several authors, gas prices dynamics at the wholesale level in Europe are explained by changes in supply and demand factors Figure 6. Those factors are related to commodity economic theory or still indexed to oil prices. The latter, are less impacted by the demand variables, such as the weather and economic factors in the consuming region, and more impacted by the main oil exporters, geopolitical considerations and speculations.

[5] These producers can withhold production to increase downstream prices for greater profits.

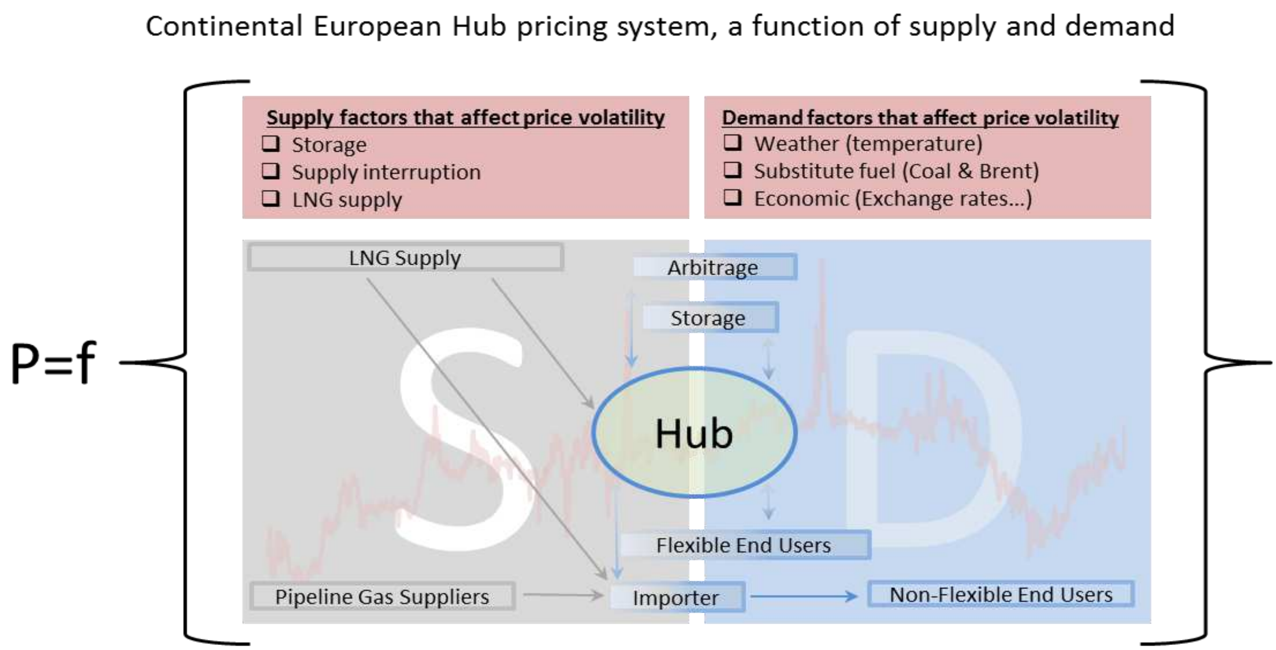

Figure 6. European wholesale market

A brief review of authors that have studied the demand and supply variables that affect the European natural gas prices is listed below:

In fact, [44] analyzes the elasticity of residential gas demand in 12 European countries. By employing a dynamic demand model, the author found that the demand is very inelastic to both price and income in the short run for all countries. [45] focused on industrial demand for natural gas in a study that spans 11 different industries in 13 European countries. Based on the results, the authors conclude that gas demand is highly price inelastic in the short run across all industries. The analysis above is in line with the fact that natural gas demand is inelastic.

Weather information reaches the market on a highly frequent daily basis and temperature changes tend to cause short-term natural gas demand variations [46]. Weather fluctuations influence natural gas prices in North West of Europe, as gas is the primary fuel used for heating.

[47] suggests that despite almost the most severe possible test of resilience demand in the UK gas system was not impacted by the outage of the rough storage system. However there was, literally, a price to pay, because gas prices moved significantly above what otherwise would have been, showing a much greater sensitivity to temperature.

There are certainly similar price trends for the following three commodities: gas, oil and coal [18]. First, there are areas where oil/coal and gas are substitutes. In the power sector, oil and coal are still the fuel of choice since oil/coal peaking units still exist. Thus, a high oil/coal price and/or high CO_2 prices would increase the incentive to achieve higher utilization of existing gas-fired power generation capacity, and consequently drive up gas demand. Second, there is a significant horizontal integration between the oil and gas industry as almost all major oil companies are also active in the natural gas business. Third and most importantly, Germany imports gas via long- term contracts that are oil price indexed.

There is also an empirical evidence for a correlation between oil/gas prices and exchange rates, i.e. an increase in oil/gas prices is accompanied by a falling dollar. The USD is a key driver for global LNG prices typically priced in USD, and given Europe's rising import dependency; these prices are increasingly influential in determining European hub prices. As a result, movements in the USD are an important factor in determining European LNG prices. If the EUR depreciates against the USD, it will erode Europe's ability to compete for available LNG in the global market (particularly against Asia). The USD influence also comes through the oil indexation of long-term gas contracts.

Gas supply contracts are either flat or flexible [48]. The first category is relatively price inelastic while the flexible supply can respond to changes in hub pricing, with flows based on the relationship between hub prices and contract prices, or an opportunity cost alternative in the case of storage. In a competitive commodity market where the demand is highly variable, storage is crucial in balancing demand and supply conditions [46].

Physically there are some storage facilities more dedicated to short-term variation in demand and some more in charge with the seasonal swing. One can reasonably say that flexible storages values will capture the correlation and causality between the prices on the spot market and storage behavior.

From 2006 onwards, disruptions, affecting the European economy and the smooth natural gas flow, were caused mostly due to non-European geopolitical, interstate conflicts or curtailment, as pointed out in [49].

Providing an additional source for gas, LNG terminals also play an increasingly important role in gas markets in the EU, a highly import-dependent region. The EU's 28 LNG imports serve as an indicator of supply conditions and might contribute to the spot prices of NCG hub

The results of the models used in this study, to forecast the gas prices, will help us assess the efficiency of such econometric models in a first step, and help us identify if the supply and demand variables listed here within, are really affecting the gas prices in a second step. More information is found in chapter 6.

Challenge 4 – The role of gas regulators: Assessing the need for further market intervention

The liberalization of energy markets is an essential and fundamental policy tool used by authorities to regulate the natural gas sector. The role of the regulators is to promote competitive conduct, domestic gas production, third-party access, price trade reporting, and ensure the presence of futures trading. Once the measures are initiated, the status of the gas hub will be confirmed as liquid and stable, and the prices are considered as indicative of market fundamentals.

Studies confirm that parameters such as market participants, the monthly day ahead trades, and churn ratio give an indication and a feel of the market. The churn ratio, calculated as the ratio of traded gas volumes to the total gas demanded, is an additional indicative measure of the liquidity of a gas hub and market maturity, and it measures the confidence of traders and consumers in the market.

In this challenge, I focus on the North American and European markets, in specific the United Kingdom (UK), since both attempted to liberalize the gas markets, and underwent intense regulations and policy changes over the past years.

The following studies [20,29,30,50–52], indicate a big difference between the volumes of gas traded in the future and the volume traded on the physical in both markets. This implies excessive participation for traders and financial players in the virtual market. Additionally, high churn ratios, indicate high liquidity and healthy trading platform, an attractive characteristic for all stakeholders. This simply means that gas suppliers in the U.S. and the E.U. markets, although in asymmetric proportions, are working in liberalized and competitive market. Although some authors, such as [53], clearly indicate that the gas suppliers in North Western Europe are oligopolistic in nature and sometimes can abuse prices, which is clearly not in favor of customer welfare - a typical example where the actions of a regulator are needed.

In the quest to validate the opinion of the cited authors/ studies, and to answer challenge four, the Information Theory will be used and further discussed in chapters 2 and 7.

1.4 Thesis outline

The aforementioned research topics are documented in several topical chapters.

Chapter 2 discusses the available modeling framework and econometric models with regard to their suitability to the natural gas market.

Chapter 3 presents the state of the art of the methodologies and the contribution of this research to the literature, more specifically in the domain of forecasting.

In chapter 4, the first modeling framework is presented and applied in order to answer our first research question related to the one global nominal price for natural gas.

Chapter 5 documents the application of another suitable modeling framework to study the effects of market concentration on gas prices.

Chapter 6 gives a comprehensive answer to what extent historical prices in one of the emerging gas hubs in Europe can be explained, and which forecasting technique is more suitable in the short and long run.

In Chapter 7, the Information Theory is used to extract important signals from gas prices that can be essential in improving the role of regulators in the relevant gas markets.

Chapter 8 lists the main findings, takeaways of the various research questions, and gives an outlook of future research directions.

Chapter 2

2. Introduction review: Mathematical models used in gas markets

This section will focus on the mathematical and econometric models found in the literature. A detailed overview is presented and listed in four different statistical categories and linked to the relevant challenges presented earlier: tail distribution models, whole distribution models, game theory econometric models, and Information Theory.

2.1 Challenge 1 – Econometric models that focus on the tail of distribution

Natural gas, traded on the wholesale market, exhibits particularly large increases in price volatility. The rise of competition and deregulation leads to relatively free energy markets, which are characterized by high price shifts. Therefore, the market is, "vulnerable" to price spikes/drops. In response to an unpredictable, volatile and risky environment, protection against market risk has become a necessity.

Accordingly, modeling the gas price fluctuations and implementing an effective tool for energy price risk management is important. Value at Risk (VaR) has become a popular risk measure in the financial industry. Since VaR estimations are only related to the tails of a probability distribution, Extreme Value Theory (EVT) may prove particularly effective.

The study of extremes focuses on outliers, a characteristic that enables a better prediction of unexpected extreme changes. The inherent stylized facts exhibited by commodity markets make the direct use of EVT complicated. For this reason, most of the applications in commodity markets involve a two steps conditional approach introduced in [54], known as the GARCHEVT approach.

The first step captures the stochastic volatility of the time series. The second step consists of applying EVT to the pseudo-independent and identically distributed (I.I.D) innovations obtained in the first step [55–58].

Another line of research includes other stylized facts, such as long-term memory, change of regimen in volatility and asymmetric effects [59]. For instance, [60] analyzes the regimen changes on volatilities for crude oil markets (Brent and WTI) and stock markets of UK, France and Japan, and finds two possible volatility regimens. [61] also considers volatility models including long-range memory for estimating risk measures for some major crude oil and gas commodities. The research showed that, models with long-range memory and asymmetry perform best in one-day-ahead forecasting. [62] explores the relevance of asymmetry and long-term memory to model. A forecast of the conditional volatility was applied in four widely traded commodities (crude oil, natural gas, gold, and silver). The findings show that nonlinear GARCH models, capturing these stylized facts, perform better in terms of volatility forecasting.

The study of extremes can also find the pair wise dependence (co-movement) between different markets that can vary from almost independent to strongly dependent in contrast to previous literature [63,64].

Several researchers have used the Extreme Value Theory in oil and commodity prices [57,58,65], however, the application of records theory to study extreme events instead of the classical EVT generates many advantages because the EVT is asymptotic, non-exact, and dependent on the choice of distribution. The majority of the findings and results of the Records Theory are exact and non-asymptotic [66,67].

In addition, several record models properties are distribution free (i.e. independent of the choice of the underlying distribution). This helps practitioners to overcome the theoretical complexity, which is hidden behind the choice of the right distribution.

The EVT approach is generally applied in a context where the observations are independent and identically distributed ($i.i.d$), which is not always a good hypothesis to be considered. Moreover, to go beyond the $i.i.d$ case in EVT makes the work even more complicated. However, record models beyond $i.i.d$ context are easily manipulated. Worthy to note, several properties retain their distribution-free nature, which is a big advantage in practical problems.

Finally, the results in EVT concentrate on the value of the extreme events because EVT studies the standardized maxima of the observations, without taking into consideration when these extreme events took place. However, records theory focuses on the values and times of extreme events, a feature that makes the potential results analysis more accurate Record Indicators [68], particular random variables, account for the study of time in record models.

The advantages offered in Records theory will be used to answer challenge one in the first place, highlighting the overall stability of three regional gas markets, and second, assess the probability of witnessing a spike/drop in the short term gas prices in three different markets. As previously explained, there have been attempts to forecast the probability of extremes in commodity prices. Yet to the author's knowledge, this is the first time the records theory is applied.

Using such theory can help gas traders have an idea about global gas prices unpredictability.

2.2 Challenge 2 – Game theory econometric models

The question of whether liberalization has led to the convergence and integration of gas markets has attracted much interest among researchers and the scientific community. Two main mathematical subjects were used to study the effect of liberalization of gas markets in literature.

Classical mathematical regression tools and statistical tests are used to analyze the effect of supply and demand fundamentals on gas prices and the market integration.

Using empirical statistical tests on gas prices, such as co-integration [69] or the Kalman filtering technique [70], the respective authors are able to identify significant level of integration in several gas hubs in Northwestern Europe. This is in line with the recent findings of [34,44], who also argue in favor of a high level of integration in gas markets.

Other areas of mathematics related to optimization equilibrium problems, more specifically non-linear complementarity problems, were used to model gas markets and suppliers' behavior [71,72].

In addition, the authors [36–41] have used such models to study the economic behavior of key market participants on different structural changes, namely increased demand, import dependency, and supply disruptions. They made use of the Karush–Kuhn–Tucker (KKT) optimality conditions. In general, the market participants behavior in the equilibrium model can range from perfect competition to monopoly, oligopolistic or Cournot players with conjectured supply functions relative to their rivals[6]. Most of the researchers agree on the Cournot competition as the most elastic representation of today's gas market, especially in the European gas market.

The models used in these studies are of a parametric nature, which focuses on the use of empirical data and assumptions such as but not limited to cost and demand functions that yields a classical optimization problem in the context of non-linear programming. The data on cost and gas contracts are normally inaccessible to the public because of non-disclosure clauses, especially with regard to old legacy gas contracts. Additionally, most of the parametric analyses rely on statistical assumptions about underlying data. The results and conclusions can only be validated if the assumptions are correct.

To complement the work done previously and to offer an alternative to parametric modeling in the field of natural gas markets, non-parametric models are used in this research in order to study the effect of liberalization on European gas markets. The latter model is developed by some authors, namely [73–76], who contribute to the development of the theory behind the application of a test that can detect a Cournot behavior of firms competing with each other using the utility maximization objective function. As explained in this research, this kind of algorithm allows to bypass many assumptions related to cost and demand functions.

In order to answer challenge two and to add to the literature in the domain of gas market concentration, this study appears to be the first to customize the non-parametric theory and apply it to the natural gas market in an attempt to overcome the constraints of parametric models.

2.3 Challenge 3 – Econometric models that focus on the whole distribution

Classical econometric models, usually focusing on the whole distribution, have been widely used in literature, especially for oil and gas prices volatility modeling. Uni-variate (mainly gas prices) and multi-variate generalized autoregressive conditional heteroscedasticity GARCH methods have been used for analyzing high-frequency time series on a data set that range from natural gas/ oil prices to supply variables, such as storage and LNG supply, and demand variables, such as weather and competent subsidies prices [46,77,78].

[6] These producers can withhold production to increase downstream prices for greater profits.

As widely agreed in the literature, inferences that do not take into consideration the regime switching phenomenon may lead to unreliable results for much high-frequency time series. In the case of oil and gas markets, sudden short period shocks will not be accounted for. Authors such as [79–81], have shown that evidence of regime-switching shall not be ignored in the behavior of natural gas prices, and that the regime switching model performs noticeably better than non-switching models in these cases.

To find evidence of causality between supply and demand variables that affect the natural gas prices, stochastic models, such as multivariate vector autoregressive (VAR) and vector error correction models (VECM) were used by [79] and [82]. Others, such as [27], study the relationship between international gas market prices and their relation to the oil price through principal components analysis and Johansen likelihood-based co-integration procedure.

In contrast to VAR and VECM models that both assume a stable relationship the relationship between the variables could be different in the separate regimes. Therefore, authors, such as [83] have used the Markov-switching vector autoregressive (MS-VAR) models. However, all the classical models cited above contain a large number of parameters, a fact that poses estimation challenges, and over-parameterization concerns [84,85].

Non-parametric and non-linear models are also used in literature. The machine learning is essential and can be used to model the complex non-linear relationship between different variables because they are “constraint-free”. Therefore, there is no need for additional tests (i.e. normality test for residuals, autocorrelation, etc.). [86] uses machine learning techniques to forecast the movement of the day ahead natural gas spot prices. A second category, such as [87], uses the neural network to predict the daily natural gas consumption needed by gas utilities. A third category, such as [88], uses Gamma test, a mathematically proven smooth test that helps machine learning modelers choose the best input combination before calibrating and testing models, a characteristic that reduces the inputs selection uncertainty.

In an attempt to answer challenge three and to provide an efficient tool for better forecasting the gas prices, four classical parametric and non-parametric methods were used, in addition to two statistical and data calibration tests, to compute the coefficients of a multivariate causal regression analysis, by linking them to a “use cause” in the subject of natural gas pricing. As previously mentioned, some authors have used parametric regression methods to forecast gas prices; however, this is the first time that both parametric and non-parametric methods are used for the same data set with the aim to compare which method is best used to better understand what drives gas prices. In addition this study is complemented by two statistical data calibration tests.

2.4 Challenge 4 – Information theory

The authors have identified three main metrics that can signal information in the hidden structure of the price values of both hubs (One located in the U.S. and the other in Europe, for more one this, please refer to chapter 7). These metrics are based on econometric and mathematical methods, and are used to inform the regulator in each market about the following:

- Signal 1: Level of competition

- Signal 2: Market stability
- Signal 3: Volatility and uncertainty of prices

The first signal studies the degree of concentration in the two different gas markets by using Game Theory, specifically the non-parametric Nash-Cournot equilibrium test. In other words, if the test shows that traders are participating in the market by trying to maximize their profit as "the only pure" strategy, then the market is considered efficient and the likelihood of anti-competitive behavior is negligible. The mathematical literature of this model is well documented in challenge two of this section.

The second signal employs the Records Theory, which relies on the analysis of the peak observations reached in a certain period, and that exceeds the previous observations. This signal measures the degree of market stability, by calculating the probability of witnessing future peak prices. Therefore, the measure of probability is a measure of market stability and predictability. If the results point towards a tendency to score high probabilities of extreme gas prices, then the market can be characterized as unstable. The mathematical literature of this model is well documented in challenge one of this section.

The third and final signal studies the price predictability of both markets, by the use of Shannon Entropy and the measure of volatility. This is done by analyzing the variation of prices and returns and assessing the degree of uncertainty and volatility which are present in gas prices. Simply, the higher the uncertainty in prices, the higher the volatility. The Shannon Entropy is used on a time series analysis, to test the predictability power hidden in the underlying probabilistic distribution of the considered time series. A time series with a high predictability power is considered to have a high level of stability with an anticipated pattern. Several researchers have previously attempted to predict the entropy of the commodity markets (oil, more specifically Brent and West Texas Intermediate, WTI[7], and other commodities) and tried to measure the information from statistical observations [89–91].

These signals combined will inform the regulator about the functioning of the market. If the market shows signs of concentration, the likelihood of extreme prices, volatility, and uncertainty, then the regulator should intervene and use its policy enforcement power.

Since our signals are based on econometric theory and models, it is important to assess the performance of such models. Therefore, and in order to answer challenge four, a quantitative analysis that relies on Information Theory is used to compare the power and efficiency of the information generated by all signals in the two different selected markets. The market with the highest information power will give additional credibility to the signals so that the price signal speaks for itself. Regulators in such a market have higher confidence and can trust the signals, which will guide their decision of whether to intervene in the market or not.

The Information Theory introduced by [92] is a probabilistic principle that helps to quantify the information generated by a random variable in an uncertain context. Information Theory can explain observations without the need to rely on neither statistical assumptions regarding the distribution of random variables, nor the random noise [93], nor the return and cost functions. The mathematical tool that will be used in this work to measure the amount of information from the gas wholesale clearing prices is the statistical entropy.

[7] Brent and WTI are two different crude oil grades (quality) and are known to be the most important oil pricing benchmark around the globe.

Another tool to assess the information in a decision-making problem is the Blackwell approach [94]. However, this approach has several complexities that prevent a simple application of Blackwell's principle [95,96], especially at the level of cost and return function assumptions. Hence, to overcome all the mentioned difficulties, the entropy principle will be applied.

By answering all four challenges, the research will analyze the effects of liberalization in the different gas markets, and check whether liberalization has contributed to less concentrate markets (where various suppliers compete), increased wholesale price correlation between the various markets (where prices react in a similar manner to changes in supply and demand fundamentals), and increased integration between markets (where the gap between the gas prices in different gas prices is reduced). The study will inspect the performance of the different new econometric models applied to gas markets, and check the gas markets forecasting performance both from a statistical and mathematical. Finally, the multidisciplinary econometric models have been combined to create a probabilistic structure, which is helpful in assessing the performance of the gas markets. This assessment is helpful to all stakeholders that participate in the gas value chain, most importantly, the regulators who play an important role in market oversight.

Chapter 3

3. Methodologies applied and progress beyond state of the art

A very rich field of modeling and forecasting has grown over time in the scientific field (see Chapter 2), where the application of statistical methods to commodity markets and prices has been limited, as opposed to its practical application. Unpredictable behavior of prices and the low power of forecasting and consistency of the mathematical models could be possible explanations for this limitation.

In an attempt to increase the understanding of the unpredictable behavior of gas prices, and in order to answer the four challenges, I will present in chapters 4, 5, 6 and 7 the details of the proposed methodologies and will illustrate the findings and the results.

Price fluctuations vary in time and can be categorized in three different time periods: Long, medium, and short term (Figure 7)

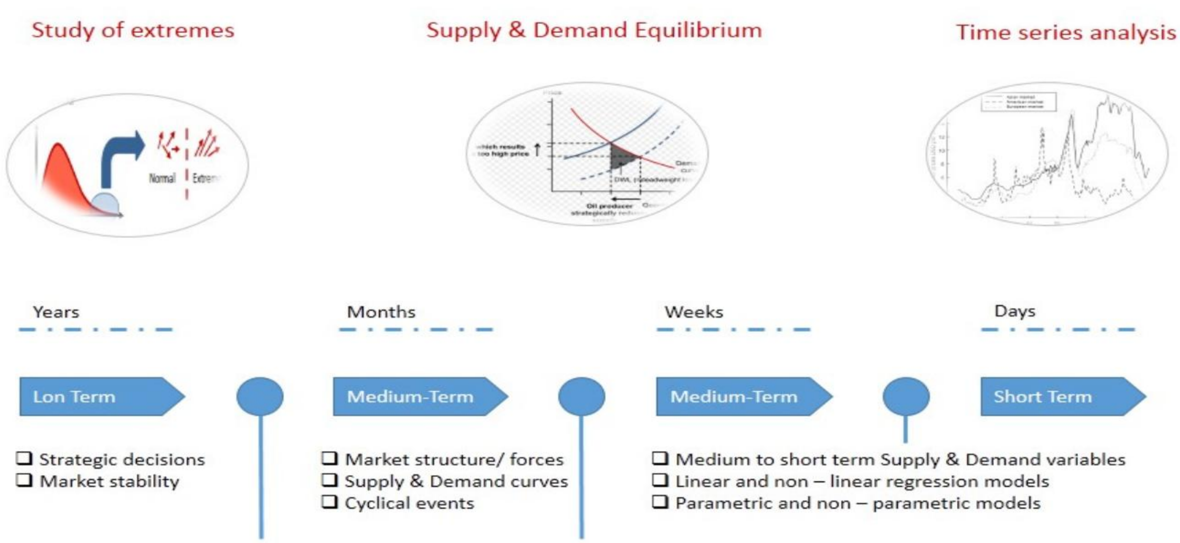

Figure 7. Three different types of commodity forecasting

3.1 Long term forecasting

Commodity markets are subject to shocks or trend changes that range from geopolitical events to structural market variations. The price shocks can be sudden, quick to mean-revert, or on other times, prices trend take a whole different level, usually to higher prices and even have chaotic behavior such as random walk. Structural market changes are mainly attributed to the market rules and regulations and the liberalization process of the market itself. As previously discussed in section 1.1, there exist three regional gas markets; each one is clustered by a regional spot market that varies in dynamics from the others.

It is almost challenging or impossible to accurately predict gas prices, especially for gas that is traded on the wholesale market, and exhibits particularly large increase in price volatility. Since forecasting real prices has a wide range of uncertainty in the long term, it is wiser to focus on computing the probability of witnessing extreme gas prices. This will support and give scientific evidence for gas market stability. Statistically speaking, the study of extremes focuses on outliers, a characteristic that enables a better prediction of unexpected extreme changes.

A first and classical method in the Records Theory is to study the behavior of record series and consider the case where the observations are independent and identically distributed (*i.i.d.).* A higher level of complexity is to go beyond the *i.i.d.* case. To do so, two possible options exist: the Yang model or the random walk model. The choice between these models will be based on statistical goodness of fit tests depending on the structure of the underlying observations.

- **Classical model**

An observation X_t is called an upper record if it is higher than all previous observations. In the case, where X_t is independent and identically distributed, $i.i.d$, the classical Record model is used, where the records are concentrated among the first observations.

- **Yang model**

In the other case, where the sequence $\{X_t, t \geq 1\}$ of independent but not identically distributed random variables is considered, the Yang model is used. For this model, the chance of having new record always exists even in the long run forecast, and is usually encountered when the analyzed variable is more or less unstable (in other economic terms, volatile). In addition, the inter-record time, i.e. the time recorded between two consecutive records, the n^{th} and the $(n+1)^{th}$, is asymptotically geometrically distributed.

- **Discrete-time Random Walk model (DTRW)**

If the observations are dependent and not identically distributed then DTRW is used.

The analytical framework that is based on distinctive mathematical models will be described in chapters 4 and 7 to answer challenge one and part of challenge four.

3.2 Medium term forecasting

Factors that affect gas prices in the medium term tend to be related to market structure and forces, more specifically in the supply and demand curves and equilibrium. The statistical models that can be used include modeling theories, e.g. Optimization, programming, and general equilibrium models (Where a producer maximizes his/her own profits given the aggregate demand function for the commodity of interest, gas in this case, and the supply response of the other firms in the industry)

Applications to world energy markets have not been as extensive because of the difficulties in dealing with regulatory and non-competitive influences on market behavior. Each market has its own regulatory regime and it can vary from a fully liberalized market to a state regulated market i.e. from one extreme to the other. The focus will be the European gas market that has evolved rapidly over the years[8], more specifically Austria and the Netherlands, as each represents a different evolutionary stage in the process of wholesale gas markets liberalization.

A comprehensive analysis which goes beyond the state of the art, is the use of **non-parametric models** in order to study the effect of gas market liberalization. This new method offers an alternative to parametric modeling in the field of natural gas markets. Besides, the application of the latter method also allows us to test for evidence of market integration.

The analytical framework is further described in chapter 5 and 7 to answer challenge two and part of challenge four.

3.3 Medium to short term forecasting

In this case, variances in wholesale commodity prices come primarily from financial markets and from short-term supply and demand variables, such as weather and storage.

Most of the econometric models and tests that investigate linear and nonlinear price fluctuations are based on mean reversion and variance measures. This form of model will be an econometric time series equation or a set of equations, matrices. The latter can be uni-variate in which a single variable is explained in terms of its past statistical history or multi-variate in which the past statistical history of several variables is combined.

By considering the multivariate model instead of uni-variate, more independent variables to explain the behavior of the dependent variable are added. This implies a reduction of the model bias derived from the error component of any econometric analysis.

Using spot prices and weekly data for other supply and demand variables of the German gas hub, the methods of least square, maximum likelihood, machine learning gradient decent and least square optimization are used to compute the coefficients of a multivariate causal regression analysis. This study also tests the short-term prediction of wholesale natural gas prices for each method used. There are four different estimation techniques that are used in this study, divided into parametric and non-parametric estimation techniques.

- **<u>Parametric methods: Ordinary Least Square (OLS) and Maximum Likelihood (MLE)</u>**

The estimation procedures of the considered parametric econometric models are the two most popular and most applied in statistics, known with the good quality of the generated estimators on bias and standard deviation levels: Ordinary Least Square classical method and the Maximum Likelihood method.

[8] This will be explained in details in chapter 4

Before considering any higher level of complexity, a simple starting point was to apply the classical OLS-Multi Linear Regression. This method is bounded by several assumptions, such as linearity, full rank (non-linear relationship), exogeneity, homoscedasticity, and normality of the error.

Another powerful estimation method is the MLE, shown to perform better and simpler to manipulate than the OLS. Another advantage is that the MLE can separate between exogenous and endogenous variables, and this generates an accurate understanding of what drives the gas prices in the short and long term. The most known models where the MLE is considered are the VAR and the GLM models.

Both models chosen above are considered to be parametric and therefore constrained by assumptions, consequently they need additional tests such as normality test for residuals, autocorrelation, etc. Two additional methods that are non-parametric are also used and are mainly based on machine learning.

- **Non-parametric methods: Gradient Decent and Least square optimization Neural Network**

Two tests are introduced to simplify the data set, reduce the number of observations, and avoid over fitting because the data set is considered large.

The first test is the Principal Component Analysis, PCA, which uses the orthogonal linear transformation that transforms the data to a new coordinate system, representing the new principal components. The aim is to get rid and omit the axis that holds the least amount of information, thus reducing the number of variables.

The step that will follow is to use the Gamma test to decrease the number of observations and data points for each variable.

The non-parametric method used is the Neural Network, which does not assume a particular family of distribution, nor is the method constrained by any assumptions. In this section two different supervised learning techniques will be used, the Gradient decent and the least square optimization. The latter can be trained in one stage rather than using the iterative process as in gradient decent, and is also good at modeling nonlinear data.

Rational mathematical and economic interpretations presented in chapter six will contribute to the understanding of what causes natural gas price volatility in the German Net Connect (NCG) hub. In addition, the comparison of four different regression estimation techniques for multivariate causal analysis has not been conducted previously in literature.

The analytical framework is further described in chapter 6, in the quest to answer challenge three.

Chapter 4

4. Modeling the price dynamics of three different gas markets – Records Theory – Challenge 1

4.1 Nomenclature

$(\Omega, \mathcal{F}, \mathbb{P})$	…	Probability space
X	…	Real random variable
Y	…	$i.i.d$ real random variables
$F(\cdot)$	…	Cumulative Density Function, CDF
$f(\cdot)$	…	Density function
t	…	Time t, equivalent to one month
T	…	Number of available observations
γ	…	Yang model parameter that needs to be computed
$i.i.d$	…	Independent and identically distributed random variables
X_t	…	An upper record if it is higher than all previous observations
$\{R_n\}_{1\le n\le N_T}$	…	Value of the n^{th} record, where n is a positive integer
$\{L_n\}_{1\le n\le N_T}$	…	Occurrence index of the n^{th} record
N_T	…	Number of records in a time series of length T
$\{\delta_t\}_{1\le t\le T}$	…	Sequence of record indicator, which is equal to 1 if X_t is a record and zero otherwise
N_T	…	Number of records in a time series of length T. In other words $N_T = \sum_{t=1}^{T} \delta_t$
ρ_t	…	Non-random integer number
$\mathbb{P}[\delta_t = 1] = \mathrm{P_t}$	…	Probability that the t^{th} observation X_t , is a record
$\mathbb{E}[\,]$	…	Expected value
$\mathbb{V}[\,]$	…	Variance
$log\, L(\,)$	…	Log likelihood function
$\mathcal{N}_T$	…	Static of a distribution
α	…	Confidence level
$N(0,1)$	…	Normal distribution with mean equal to zero and variance equal to 1
Δ_{L_n}	…	As the inter-record time: Time recorded between two consecutive records n^{th} and the $(n+1)^{th}$
η_t	…	$i.i.d$ increments drawn from a continuous distribution (error)

Table 2. Notation for the models applied in chapter 4

4.2 Introduction

As explained in chapter 1, there are three different regional gas markets: The US, the European and the Asian markets. In Europe and the US's regional market, natural gas is mostly purchased through pipelines due to large domestic resources and strong grids. The lack of such infrastructures in North East Asia prevents the import of natural gas through pipelines. Therefore, natural gas could be only imported in the form of Liquefied Natural Gas (LNG), which is shipped on maritime tankers. The highest demand for natural gas in 2017 was in North America, a value that is closely followed by Europe, then Asia.

Two basic pricing systems are commonly used for international trade of natural gas. The split in price formation varies deeply between regional markets, depending on several structural factors such as regulation, liberalization process, contracting practices, existence of a spot market, liquidity, and share of imports.

- Gas-on-gas pricing, where the price of natural gas is competitively determined based on gas market spot prices. As such, prices vary as a response to natural gas supply and demand.

- Oil-indexation pricing, where the price of natural gas is determined based on oil market spot prices. As such, prices vary as a response to oil supply and demand.

It is almost challenging to get an accurate assessment of gas indexation levels and growth. Most of the supply contracts are structural long term portfolio contracts which are considered to be highly sensitive confidential information. Typically contract terms and conditions have evolved over time through the processes of renegotiation and price re-openers

Facts have shown that natural gas, which is traded on the wholesale market, exhibits particularly large increases in price volatility. The rise of competition and deregulation leads to relatively free energy markets, which are characterized by high price shifts. Therefore, the market is, "vulnerable" to price spikes/drops. In response to an unpredictable, volatile and risky environment, protection against market risk has become a necessity.

The objectives of this research is, first to study the stability of the three regional gas markets by using the records theory, and second, to assess the probability of witnessing a spike/drop in the short term gas prices in three different markets. As previously explained, there have been attempts to forecast the probability of extremes in commodity prices. Yet to the author's knowledge, this is the first time in which the records theory is applied.

Since the time series of the markets are mostly $\mathrm{non-i.i.d}$, and the assumption of the type of distribution is complex, the best approach to find a feasible solution for the gas markets is to use the records theory. The models will be tested to check the reliability of the results. Testing is done by comparing the theoretically expected number of records to the real ones. This will help relevant stakeholders to better estimate their risk portfolio in the short term.

In section 4.3, the most popular record models will be presented. In section 4.4, the significance of the results and its impact will be explored.

4.3 Material and methods

4.3.1 Study area and data used

The description of the time series of the three different gas markets that will be used in our study is listed in Table 3. All prices are denominated in USD per MMBtu (millions of British thermal units). The data set consists of monthly values recorded between $March\ 1997\ to\ November\ 2016$.

Variable	Frequency	Number of observations	Description	Unit	Source
JCC	Monthly	237	The Japan Customs-cleared Crude prices (1997-2016)	$USD\ per\ MMBtu$	PAJ[9]
HH	Monthly	237	The Henry Hub prices (1997-2016)	$USD\ per\ MMBtu$	U.S. Energy Information Administration, EIA [10]
GBP	Monthly	237	The German Border Prices (1997-2016)	$USD\ per\ MMBtu$	European Energy Exchange, EEX[11]

Table 3. Data collection - Records Theory

The data is rich in statistical parameters that could be interpreted using statistics (such as correlation, trends, stationary, etc.) and economics tools.

4.3.2 Record theory models

In this document $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. X a real random variable defined on Ω with a cumulative distribution function (*CDF*) $F(\cdot)$, and a density function $f(\cdot)$. It is supposed that $(\Omega, \mathcal{F}, \mathbb{P})$ is sufficiently rich to support an infinite sequence $\{X_t, t \geq 1\}$ of independent copies of X; they are therefore independent and identically distributed $(i.i.d)$ random variables. An observation X_t is called an upper record if it is higher than all previous observations.

[9] Available at the Petroleum Association of Japan, http://www.paj.gr.jp/statis/statis/data/07/paj-7_201701.xls
[10] Available at the Energy Information Administration, https://www.eia.gov/dnav/ng/hist/rngwhhdm.htm
[11] Available upon request from the European Energy Exchange AG, marketdata@eex.com

The value and the occurrence index of the n^{th} record are respectively given by the following sequences $\{R_n\}_{1\leq n\leq N_T}$and $\{L_n\}_{1\leq n\leq N_T}$. Such that $R_n = X_{L_n}$ and N_T is the number of records in a time series of length T. In addition, the sequence of record indicators $\{\delta_t\}_{1\leq t\leq T}$ is defined and is equal to one if X_t is a record and zero otherwise. Therefore, it is easy to see that $N_T = \sum_{t=1}^{T}\delta_t$.

4.3.3 Independent and identically distributed (*i.i.d*) case

Record's properties in the case where X_t are $i.i.d$ was well studied by many authors [66,67]. It turns out that many of these properties are distribution-free, i.e. they hold for any distribution of X_t. This has led to a huge progress in the overall understanding of the stochastic behavior of records. I present some of these results in order to give an overview of some important facts in the theory of records in the $i.i.d$ case.

$\mathbb{P}[\delta_t = 1]$ is denoted by $\mathrm{P_t}$ the record rate at time t. It is the probability that the t^{th} observation X_t , is a record. [67] Shows, that for all $T \geq 1$, the random variables $\delta_1, \dots, \delta_T$ and $M_T = \max(X_1, \dots, X_T)$ are mutually independent with $\delta_t \sim Bernoulli\left(\frac{1}{t}\right)$, so that

$$P_t = 1/t \tag{1}$$

Note that the record rate goes asymptotically to zero, i.e. records are concentrated among the first observations. Based on the previous reasoning and by writing $N_T = \delta_1 + \cdots + \delta_T$ I obtain the expected value of N_T:

$$\mathbb{E}[N_T] = \sum_{t=1}^{T} 1/t \tag{2}$$

4.3.4 Non - Independent and identically distributed case

[66] have shown, in the $i.i.d$ case, that when t increases, records tend to become more spaced over time. However, in many real data sets, this phenomenon is not true for many reasons [97]. As a result, more comprehensive models were developed, where the number of records grows faster than in the $i.i.d$ case, and where records are not only concentrated among the first observations as in the $i.i.d$ classical model.

The most popular models beyond the $i.i.d$ case are the Linear Drift Model (LDM) and the Yang model where the observations are independent but not identically distributed. On the other hand, another popular model is the Discrete-time random walk, where the observations are dependent and not identically distributed.

Before considering a higher level of complexity and going beyond the $i.i.d$ model, one should test if the data fits or not the $i.i.d$ case. One option is to test the null hypothesis that the data comes from a sequence of $i.i.d$ random variables. This goodness-of-fit test can be based on the result of [66] who showed that under the null hypothesis the statistic

$$\mathcal{N}_T = (N_T - log\, T)/\sqrt{log\, T} \tag{3}$$

converges to a standard normal distribution denoted by $N(0,1)$.

Thus, one rejects the $i.i.d$ case if $\mathcal{N}_T$ is greater than the theoretical $(1-a)^{th}$ quantile of the standard normal distribution (α is the confidence level of the statistic $\mathcal{N}_T$ generally fixed to 5%). If the null hypothesis is not accepted, then the classical $i.i.d$ record model can no longer be considered, and other models that go beyond the classical models should be adopted.

4.3.4.1 Yang record model

The $Yang - Nevzorov$ model was introduced by [98]; it was shown to fit several sets of data. In the latter model, a non-random integer numberρ_t, of $i.i.d$ random variables Y with a cumulative distribution function $F(\cdot)$, is generated and available simultaneously at time t, from which is extracted $X_t = \max(y_1, y_2, \dots y_{\rho t})$.

Thus, the sequence $\{X_t, t \geq 1\}$ of independent but not identically distributed random variables is considered, with the following cumulative distribution function

$$F_{X_t}(x) = F(x)^{\rho_t}, \rho_t > 0. \tag{4}$$

Due to the $i.i.d$ property of the underlying random variable Y, the probability of a record among the newly generated $\rho_t$variables is given by:

$$\mathbb{P}[\delta_t = 1] = P_t = \rho_t / S_t \, , t \geq 1, \tag{5}$$

where

$$S_t = \sum_{k=1}^{t} \rho_k. \tag{6}$$

[99] shows that, in general, the expression of the previous probability holds if ρ_t are real and strictly positive. Moreover, he shows that the independence of the record indicators $\{\delta_t, t \geq 1\}$ remains valid for any underlying distribution. So, the sequence of δ_t is a Bernoulli process with probability of success $\mathrm{P_t}$. Note that if $\rho_t = 1 \; \forall t \geq 1$, the Yang-Nevzorov model is simply the classical $i.i.d$ model.

I recall that the distribution of the number of records N_T is related to record indicators by the following relationship:

$$N_T = \sum_{t=1}^{T} \delta_t \tag{7}$$

Thus, the expected value and the variance of N_T are:

$$\mathbb{E}[N_T] = \sum_{t=1}^{T} P_t \; and \; \mathbb{V}[N_T] = \mathbb{E}[N_T] - \sum_{t=1}^{T} P_t^2 \tag{8}$$

Now I assume the following parametric form, originally proposed by [98], $\rho_t(\gamma) = \gamma^t$with γ strictly larger than one. Thus, the probability that the X_t , is a record for the Yang model is given by:

$$\mathbb{P}[\delta_t = 1] = P_t(\gamma) = \rho_t(\gamma)/S_t(\gamma) = \gamma^t(\gamma - 1)/\gamma(\gamma^t - 1) \quad (9)$$

Note that $\rho_t(\gamma)$ represents an exponential growth in the number of available random variables at time t. Furthermore, in a Yang model the record rate goes asymptotically to a constant given by:

$$(\gamma - 1)/\gamma \quad (10)$$

This means that the chance of having a new record always exists even in the long run forecast. This case is usually encountered when the analyzed variable is more or less unstable (in other economic terms, volatile).

In order to make Yang model applicable in practice, the goal is to estimate the parameter γ, based on the maximum likelihood estimation method[12] and by using the probability distribution of the record indicators. As the record indicators δ_t are independent and follow the Bernoulli distribution of parameter $\mathrm{P_t}(\gamma)$, the work consists in finding the γ, which maximizes the Log-Likelihood function[13]:

$$log\ L(\gamma) = log\ \mathbb{P}[\delta_1, \ldots, \delta_T; \gamma] \quad (11)$$

Still in the context of a Yang model, I define:

$$\Delta_{L_n} = L_{n+1} - L_n, n \geq 1 \quad (12)$$

as the inter-record time, i.e. the time recorded between two consecutive records n^{th} and the $(n+1)^{th}$. It has been shown that the inter-record time Δ_{L_n} are asymptotically geometrically distributed.

To summarize, the Yang model distinguishes itself from other models as, first it can fit observations that are independent but not identically distributed, second, the inter-record time follows a geometric distribution and last, the record rate goes asymptotically to a constant.

For the remaining of this section, the goodness-of-fit for a Yang model is evaluated by checking whether the observed values of the inter-record times, after a warm-up period allowing the asymptotic effect to settle, are in agreement with the geometric distribution.

The most popular approach consists in constructing a goodness-of-fit test based on Pearson's chi-square test [100]. It is assumed that the event $N_T = m$ is not random and I chose a $K > 1$ in order to partition the set $\{1,2, \ldots, \infty\}$ into K subsets. Concerning the problem of choosing K, this is a question that users have asked statisticians at the very moment when the chi-square test appeared [101–105]. Few of these responses reinforce each other, while others contradict. However none appears to be universally accepted. Recently, [106] write that the problem must be solved ad hoc in relation to contextual elements as a serious and rigorous user of this test will not fail to obtain[14].

If the Pearson chi-square test rejects the geometric distribution hypothesis, this casts doubt on Yang's model. Then, a more general model where the observations are dependent and not identically distributed should be considered.

---

[12] The maximum likelihood estimation is a method of estimating the parameters of a statistical model

[13] Refer to annex A.1, for further information regarding the maximum likelihood method

[14] Refer to annex A.2, for further information regarding the goodness-of-fit test of the Yang model

The simplest way to check if the observations are independent is to use the $Ljung - Box$ test, and if equals to unity, indicates that the observations belong to a dependent set of data.

4.3.4.2 Discrete-time random walk model (DTRW)

One can think of a more general model, where the observations are dependent and not identically distributed. The most used model in this context is the Discrete-time random walk model introduced by [107].

In this model, the t^{th} observation of the upper record is given by

$$X_t = X_{t-1} + \eta_t, \tag{13}$$

Where η_t i.i.d increments are drawn from a continuous distribution.

The stochastic behavior of this model was carefully studied, and it has been shown that the record rate is given by:

$$\mathbb{P}[\delta_t = 1] = P_t = \binom{2t}{t} 2^{-2t}\,, \tag{14}$$

In this model, all the above properties are in a distribution-free context (independent from the choice of the distribution of X_t). In addition, it is remarked that for a big value of t, the record rate goes asymptotically to zero. This means that records are concentrated among the first observations.

In the next section, results of the different tests that were used in the models are presented. The analysis and choice of models is explained and validated based on economical and mathematical reasoning.

4.4 Results and discussion

4.4.1 Results of the three models

The analysis starts by extracting the record values, indices and indicators, as well as the number of records that each market has witnessed within the period of March 1997 to November 2016.

Markets	HH	GBP	JCC
Number of records	_8_	_10_	_17_
Record index	Apr-97	Apr-97	Apr-97
	May-97	Jul-97	Jul-97
	Aug-97	Oct-97	Aug-97
	Sep-97	Jul-98	Apr-98
	Dec-97	Oct-98	Jul-98
	Jun-00	Oct-99	Aug-99
	Dec-00	Jan-00	Oct-99
	Jan-06	Jan-08	Sep-04
		Jul-08	Jul-05
		Mar-09	Feb-06
			Oct-07
			Jan-08
			Aug-08
			Sep-08
			Dec-08
			Feb-09
			Apr-15

Table 4. Number of records and record index

I start by applying the goodness of fit test, which tests the null hypothesis that the data comes from a sequence of $\mathrm{i.i.d}$ random variables. As shown in Table 5, both the German and Asian market rejects the null hypothesis, which explains empirically that those markets are struggling with sudden price variations going beyond the $\mathrm{i.i.d}$ case.

This result can be expected based on the analysis of Table 4, where the American market has the lowest number of records, while the Asian market has the highest number of records. Worthy to note, the records are concentrated among the first observations in the American market, which is typical for the $\mathrm{i.i.d}$ case.

This is in line with mathematical theory, which suggests that the number of records for $\mathrm{non-i.i.d}$ case grows faster than in the $\mathrm{i.i.d}$ case.

Next, the less stable markets, where the classical models cannot be adopted, will be developed.

Markets	HH	GBP	JCC
P-value	0.139	0.026	4e-007
Result	Accept H_0	Reject H_0	Reject H_0

Table 5. Goodness of fit test results at a confidence level of 5%

Note that the assumption of a Yang model is increasingly consistent, when the p-value of the $\chi(\tilde{\gamma})$ statistic is increasingly greater than the confidence level α generally fixed at 5%. The results shown in Table 6, suggest that both markets accept the null hypothesis, which means that the Yang model can be applied to both the JCC and GBP markets. However the GBP market has a p-value which is not considerably high. Accordingly, I explored other general models where the observations are dependent and not identically distributed.

Markets	GBP	JCC
P-value	0.127	0.358
Result	Accept H_0	Accept H_0

Table 6. Pearson chi-squared test results at a confidence level of 5%

In order to test whether the observations in the GBP market come from a dependent set of data, the $Ljung-Box$ test is used, with the aim of knowing if the observations are auto-correlated. The result of the $Ljung-Box$ is equal to unity. This implies that the series of GBP belongs to a dependent set of data, which is common to see in a commodity that exhibits lots of volatility. Consequently the Random walk model, can be applied for the GBP market.

In addition to the mathematical explanation, it can be demonstrated that GBP exhibits more volatility in economic reasoning:

First, gas in continental Europe is slowly, but surely, heading towards a gas to gas pricing rule (The latest report of IGU states that over 60 % of traded gas in Europe is indexed to the continental gas hubs). Therefore, this could bring more volatility to the market.

Second, the supply portfolio in such markets is an important factor. Thus, all supply sources coming from regions that show sign of volatility, insecurity or instability could increase volatility at times where the system does not work properly (excess of gas supply, implies less volatility and vice versa).

Last but not least, the high level of storage capacity in European gas markets would likely have a strong and important effect on price volatility as it will tend to decrease, mainly because it will increase the strategic seasonal supply of gas. However, with no diversified, secured and long lasting gas supply to Europe, this could make the presence of storage capacity unexploited.

Markets	Probability of records	Probability of having a record on t = 200 (November 2013)
HH	$\mathbb{P}[\delta_t = 1] = \mathrm{P_t} = \frac{1}{t}$	$\mathbb{P}[\delta_t = 1] = 0.005$
JCC, where $\boldsymbol{\gamma}$ is equal to 1.058	$\mathbb{P}[\delta_t = 1] = \mathrm{P_t}(\gamma) = (\gamma^t(\gamma-1))/(\gamma(\gamma^{\mathrm{t}}-1))$	$\mathbb{P}[\delta_t = 1] = 0.0555$
GBP	$\mathbb{P}[\delta_t = 1] = P_t = \binom{2t}{t} 2^{-2t}$,	$\mathbb{P}[\delta_t = 1] = 0.0399$

Table 7. Probability of Records for each market

The probability of having a record can be computed for each market and for any time in the near future. Table 7, shows the result of our mathematical models. The probabilities were computed for the observation that coincides with the date of November 2013.

The probability of having a new record is highest in Asian markets. This provide evidence that the theoretical reasoning used in the study, is precise. The following are the main outcomes:

1) The probability of having a record is low for all markets. Unless a major rupture in supply and demand fundamentals occurs, the probability of a major spike/drop is not significant.

2) Viewed as increasingly scarce a decade ago, when its price rose above $15 per million Btu (Figure 3), the substantial shale gas exploitation in the US, led to a resource surplus and substantial low prices. The result of the classical model that is applied to the $i.i.d$ case is in line with this economic fact, and the probability of having a record tends to zero in the American Market (HH). This means that there is little chance in encountering a price/drop in gas prices in the short term, a result also endorsed by recent literature [28].

3) There is always a certain probability of having a record in the European market and the Asian market. Again, the Yang and Discrete-time random walk models performed well.

   This is in line with economic reasoning for both markets.

   Research papers published by [20,30], have increasingly observed that the move from oil-indexed to hub or market pricing is a clear secular trend, strongest in northwest Europe and spreading southwards and eastwards. The strong gas markets integration between the different European hubs and, the diversification in gas supply will keep the gas prices resistant to sudden spikes/ drops. This is the result of many years of pro-competition EU regulatory initiatives implied by the several EU directives, namely the legislation concerning security of supply at EU level (994/2010). The annual surveys on pricing of wholesale gas undertaken by the International Gas Union also offer quantitative evidence of these trends [5].

On the other hand, Asian gas prices are still dominated by oil indexation of LNG contracts, which cannot follow gas market fundamentals in a timely manner. This economic reasoning is in line with recent publications, particularly for the Asian market. Research showed, that the Asian Pacific LNG is currently undergoing considerable change and uncertainty, and the risk of unexpected oil price shocks has always been the main factor in analyzing LNG trade and market interactions in the Asian gas market [21].

4) Results suggest that the Yang model fit the JCC market, and that the result of the maximum likelihood principle based on record indicators (described in Section 4.3.4.1) gives a value for γ to be equal to1.058. As γ exceeds unity, the choice for the Yang model is reasonable [11,28,108,109].

5) The record rate in some of the models converges to a certain constant in the short run. This is a significant indicator showing that the markets are in an unstable situation and vulnerable to future spikes/ drops.

4.4.2 Difference between empirical and theoretical findings

The analysis carried in this study, started by applying several goodness-of-fit tests in order to assign each set of data to the model that fits observations the most.

In a second step, the probability of the records for each model was computed.

In a final step, in order to test the models, the theoretically expected number of records for each model will be computed, based on the expected value of N_T[15], and will be compared to the real number of records extracted directly from the series of observations.

By completing all three steps, the forecasting performance of each model will be evaluated. The closeness between the theoretical and empirical results is a proof of good performance.

Markets	HH	GBP	JCC
Number of records (Actual)	8	10	17
Number of records (Theoretical)	6.043	16.36	16.47
Percentage error (%)	24	63	3

Table 8. Results of empirical and theoretical probabilities

[15] Refer to equations numbers seven and eight

As shown in Table 8, the percentage error is low for both the HH and JCC markets. This is perceived as a good indicator regarding the choice of our models: they fit the considered data sets.

The percentage error in the JCC market is minimal, which proves that the Yang model has successfully predicted the number of records in the last twenty years. Since the record rate in a Yang model goes asymptotically to a constant, this is further proof that the Asian market is unstable in the short and long run prediction and will always be vulnerable to spikes and drops in natural gas prices.

The gas regulator in the relevant Asian countries can benefit from the theoretical findings as basis to develop, support and improve business environment for developing a better functioning LNG market. The scientific and political community have shared policy recommendation to improve the situation in the Asian gas market, including:

- Liberalizing domestic gas market and developing adequate and accessible gas infrastructure capacities by promoting third party access[16] [26].
- Moving away from JCC to a price mechanism which reflects anticipated market fundamentals of the Asian buyers' country[17] [21].
- Securing new gas sources, such as East Africa and Australia to diversify the supply portfolio of Asian importers.
- Promoting governmental adequate and accessible infrastructure developments (upstream/ downstream) through: equity participation of importers/ Asian shippers in LNG projects, investment in storage in order to increase domestic flexibility, and investment in new Asian upstream projects

Consequently, the records theory on a data set that goes beyond the $\mathrm{i.i.d}$ case was successfully used and modeled without going through the conditional approach which is mostly used in EVT. In addition, the results are distribution free, which minimize the complexity of the models.

[16] The Third Party Access, TPA should be complemented by enhancing contractual flexibility. This can done by eliminating destinations restrictions in LNG contracts

[17] The imperative for buyers will be not to lock themselves into long term inflexible price arrangements during a period when market dynamics will be changing rapidly.

Chapter 5

5. Modeling post liberalized European gas market concentration– A Game theory perspective – Challenge 2

5.1 Nomenclature

i	...	Gas suppliers
t	...	Time t, equivalent to one month
T	...	Number of available observations
N	...	Number of firms in a market
$\mathrm{P_t}$	...	Gas price at time t
$Q_{i,t}$	...	Quantity of gas supplied by supplier i at period t
HHI_t	...	Herfindahl-Hirschman index, HHI at time t
$\{i,j\}$	...	Two gas firms i and j competing for market share
S_i	...	Set of strategies for firm i
$u_i(s_i,s_j)$	...	Payoff of firm i
$C(.)$	...	Cost function
a_t, b_t	...	Regression parameters
$P_t(Q_t)$	...	Market inverse demand function
α_i, β_i and γ_i	...	Cost parameters for the capacity utilization marginal cost function
$\delta_{i,t}$	...	Array of marginal cost of supplier i at period t
$\mathcal{C} = \left\{P_t, \left(Q_{i,t}\right)_{i\in\mathcal{I}}\right\}_{t\in\mathcal{T}}$	...	Set of observation that consist of a price and a quantity at time t

Table 9. Notation for the models applied in chapter 5

5.2 Introduction and literature review

5.2.1 Current state of the European gas wholesale market

The European Union (EU) gas market has rapidly evolved over the years. Several gas directives issued by the European Commission demonstrate the evolution of the liberalization process in Europe since 1998. Major developments in these regulatory reforms include providing customers and suppliers with third party access to infrastructure, a clear-cut separation in energy companies through ownership unbundling, and regulatory supervision of the member states

Combined with the exclusion of destination clauses, the adoption of the *Entry-Exit* system for capacity booking[18], in addition to the requirement to establish a *Virtual Trading Point* in a given system[19], has entirely transformed the gas market setting in Europe. Thus, trading in the wholesale markets is facilitated, and users can optimize their portfolios.[20]

Crucially, these reforms have also promoted capacity allocation tools, i.e., auction procedures for the benefit of small to medium competing shippers and traders. This was achieved by the successful establishment of the *Network Code* in the gas transmission system. As a result, congestion in the EU's gas transmission pipelines is reduced, and the efficient use of existing capacities is optimized.

Other measures[21] have also contributed to the further development of the gas markets, such as new investments in cross-border capacity, the integration of old legacy or long-term contracts into the new system, and the establishment of cross-border cooperation between Transmission System Operators via the European Network for Gas. The latter bring together network expansion and reinforcement, including cross-border capacities and interconnectors.

Thus, gas consumers in the EU have a broader choice of credible suppliers in comparison with the previously used point-to-point system. Moreover, according to [21], there is increased competition in several member states, which are mainly located in Northwestern Europe. Policy makers aim to integrate the different national wholesale markets into a bigger gas market in order to minimize any attempt by the suppliers to exert market power.

5.2.2 Gas market characteristics in Austria and the Netherlands

The functioning of the European gas wholesale markets should be measured quantitatively to reflect the gradual advances in supply-side competition, the improved price convergence across market areas, the wholesale price mechanism, the enhanced interconnection between markets and the overall integration of national markets.

The markets of Austria and the Netherlands will be the focus of the study, as it is believed that each exemplifies a diverse evolutionary phase in wholesale gas markets liberalization

Austrian Gas transmission network is composed of three market zones: Tyrol, East and Vorarlberg. The market zones of Tyrol and Vorarlberg are just associated with the German transportation network, have no physical link to Austria, and have neither storage nor indigenous gas production. The Eastern market area is where the vast majority of the Austrian gas demand is located [32]. Within the Market's East Area, the import stations are connected to the border points by the domestic distribution system and major transit pipelines exist.

[18] Moving away from predefined point-to-point transportation routes, which used to impede gas trading.
[19] This means that gas can now be traded irrespective of its location.
[20] Shippers may swap gas between locations, a factor that allows them to access gas volumes from locations to which they have no direct physical connection, avoiding the need to book and pay for or use unnecessary capacity.
[21] More information on gas reforms in Europe can be found in the Gas Directives No. 30/1998, 55/2003, 715/2009 and 984/2013.

Since 2013, within the new entry/exit model, Market's East Area formed one entry/exit zone, with one central VTP. Settlement at the VTP is carried out by Central European Gas Hub GmbH (CEGH) [110]. This has created trading possibilities on the wholesale level, which is a major shift for a market that was governed by long-term oil-indexed gas contracts. In addition, and with the adoption of the transmission network code, new traders wishing to exchange gas titles now face limited congestion in the transmission networks [110].

Austria imports more than two third of its inland consumption. Most of the gas entering Austria comes from Russia, through Gazprom, to the key entry point of Baumgarten. The remaining comes from Germany and Norway through the West-Austria-Gasleitung (WAG) pipeline, the physical connection at the border with Germany. The indigenous production is supplied by both companies OMV and RAG AG.

The Dutch gas company Gasunie is operating an entry-exit tariff system for gas transmission network in the Netherlands, similar to the Austrian one market region. The Title Transfer Facility (TTF) is the virtual point, where traders operating in Netherland, exchange gas titles [32].

Gas is exported and imported by means of connections that are located near the borders with Germany and Belgium. Gas is also imported from the United Kingdom via Belgium, through the bi-directional interconnector. Nonetheless, gas can only be imported via the connection with Norway via Emden, northern Germany [111].

There are two main qualities of gas in the Dutch system: a low calorific value (L-gas), and a high calorific value (H-gas) [112]. L-gas originates from the Groningen field and smaller fields in Netherlands and Germany.[22] On the other hand, H-gas comes from small onshore and offshore Dutch fields and from other sources such as imports from Russia, Norway and LNG. The Netherlands is the principal gas producer in the EU. More than 2/3 of total gas produced inside the country comes the Groningen field and other onshore fields. The remaining is produced in 150 fields located offshore in the North Sea [113].

5.2.3 Literature review

The amount of firms trading at a hub gives an indication on the willingness of traders to be involved and the ease of participation. More importantly, the number of active players is also indicative i.e. a larger number can be a sign of more competition, and consequently less market manipulation.

Hub	Market participants		Active participants		Total traded volumes (TWh)	
	2011	2014	2011	2014	2014	2015
TTF	60	130	45	45	13,555	17,080
CEGH	40	53	15	15	400	340

Table 10. VTP participation

[22] H-gas can be transformed to L-gas by adding nitrogen.

According to the results of [29], summarized in Table 10, TTF outperforms CEGH as it exhibits high liquidity and minimal spread. The low quantity of gas traded in CEGH is an indication of low liquidity in Austria, as compared with The Netherlands.

The traded volumes, as compared with the size of the market, determine the churn rate, which according to many references [24,29,52,69] is the main factor in determining the success of a trading platform. According to their results, shown in Table 11, TTF scores a high churn ratio, which is attractive for traders and financial players. As a consequence, participants such as banks and hedge funds can trade options on the TTF exchange. However, this is not the case of CEGH; the low churn rate and the low quantities of traded volumes is an indicator of low liquidity in Austria, an unattractive environment for traders and financial players.

Hub	Churn rates		
	2011	2014	2015
TTF	13.9	36	45.9
CEGH	2.2	4.8	3.9

Table 11. Churn ratio

Based on the presented numbers, one can directly conclude that the TTF hub can be labeled as dynamic and free from price manipulation because of the liquidity and wide range of participants in the trades on both OTC (Over the counter) and exchange. However, this is not the case in Austria.

As stated in chapter 1 and according to Table 1, the gas-on-gas competition in central Europe has increased from zero in 2005[23] to over 56% in 2015 with changes accelerating from 2013 onwards [29]. The indexation level in the TTF has reached higher norms.

This is mainly due to the liberalization process, and by 2013, a big portion of the Dutch and Norwegian long term contracts had moved to hub prices. Moreover, Russia and some of its customers agreed to revise the pricing formula, by reducing the base price by about 10%, and by reducing the take or pay commitment from as high as 90% to a value of 60% [20].

The only supply source that is still resilient to change comes from Algeria, where the gas supply contracts are still indexed on oil.

On the other hand, companies like, Statoil, Gasterra and UK producers shifted their gas contracts to hub indexation. Gazprom, Sonatrach, other key producers and several LNG exporting companies prefer long-term bilateral contracting with a higher presence of oil-price indexation [42].

The mentioned studies show that the effects of the liberation process have a positive impact in Europe. More specifically, the north western part of Europe reflects gas market supply and demand fundamentals. However, this is not yet fully the case in central part of Europe, namely the CEGH hub [25].

[23] This means that prior to 2005; the contracts in the eastern part of Europe were wholly governed by long term contracts indexed on oil.

5.3 Material and methods

In this section, an inclusive analysis on the market concentration of the wholesale gas markets in the Netherlands and in Austria will be presented, using parametric and non-parametric Game Theory analysis.

The remaining of this chapter is structured as follows: The following section will present the data and the two main methods used in the study, while section 5.4, will explore the importance of the results and pinpoint its impact.

Translated into modeling, I consider a country importing gas from i different suppliers and T the number of available observations. For each $t \in \mathcal{T} = \{1, \dots, T\}$ and $i \in \mathcal{I} = \{1, \dots, I\}$, P_t is the gas price at period t and $Q_{i,t}$ the quantity of gas supplied by supplier i at period t. The monthly data related to gas supplies are illustrated in Figure 8 and Figure 9.

The Austrian market is supplied from the east on the one hand, where a single company is the only supplier, and from Germany on the other hand, where several big gas suppliers are active. Gas consumed in the Netherlands mainly comes from local production, specifically the Groningen field, in addition to further onshore and offshore Dutch fields. The Netherlands is the principal gas producer in the EU. Other sources feed the local market, such as imports from Russia, Norway and LNG.

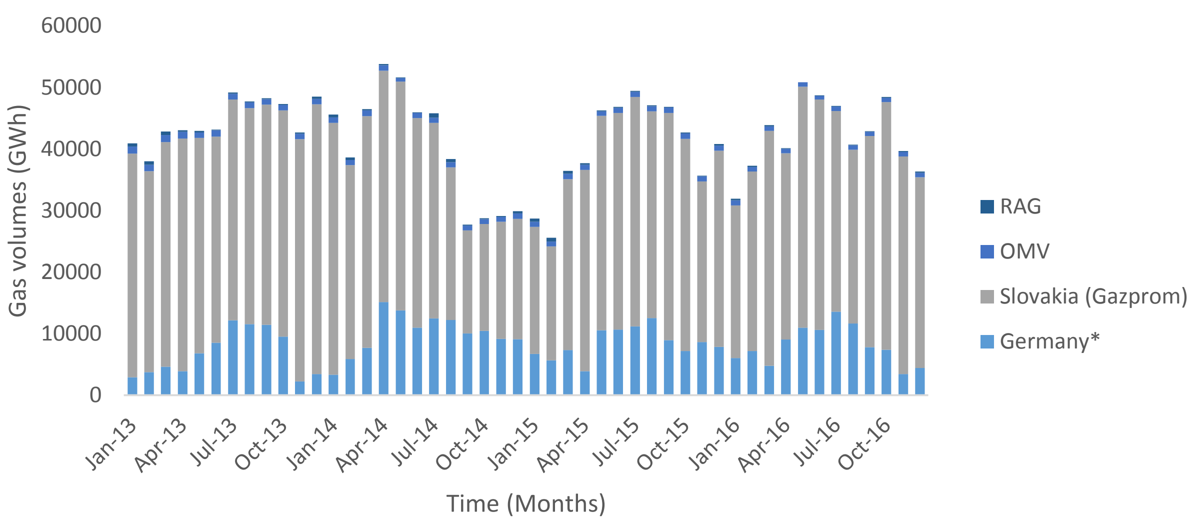

Figure 8. Monthly gas imports to Austria and indigenous production[24].

*Additional suppliers active in the market Area East receive gas from several gas shippers that are active in Germany

[24] Source: E-Control. https://www.e-control.at/en/

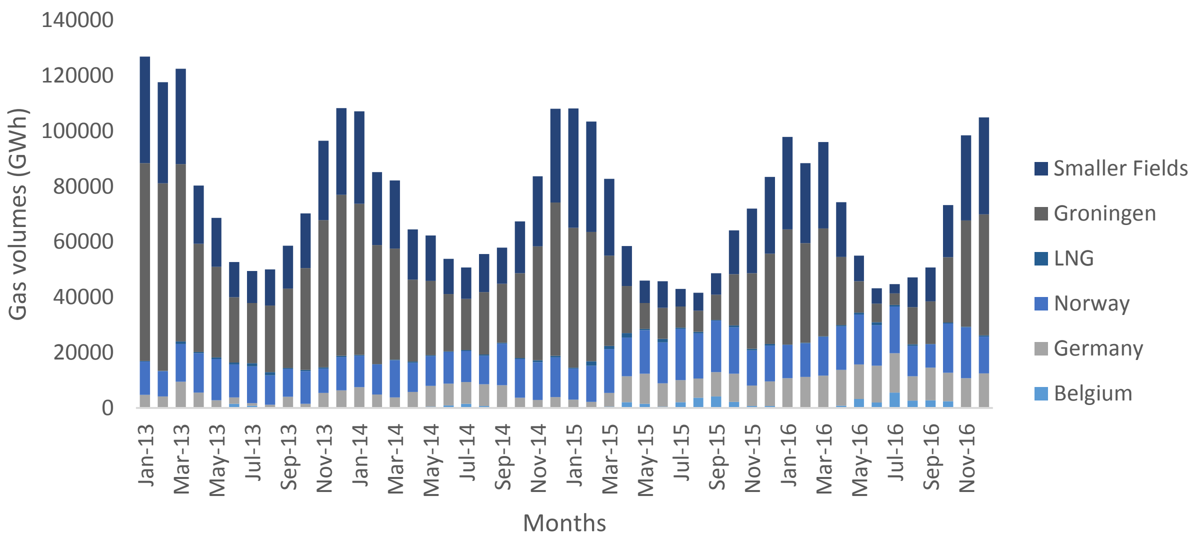

Figure 9. Monthly gas imports The Netherlands and indigenous production[25].

Figure 10 shows the time series of the natural gas prices of both the Dutch market, represented by the TTF hub, and the Austrian market, represented by the CEGH. All prices are denominated in Euros per MWh. The data set consists of monthly values recorded between January 2013 and December 2016. As shown in the graph, both price trends are to some extent positively correlated, albeit the price of gas sold on the Austrian side is at most times higher than that on the Dutch market.

[25] Source: Gasunie. https://www.gasunie.nl/en

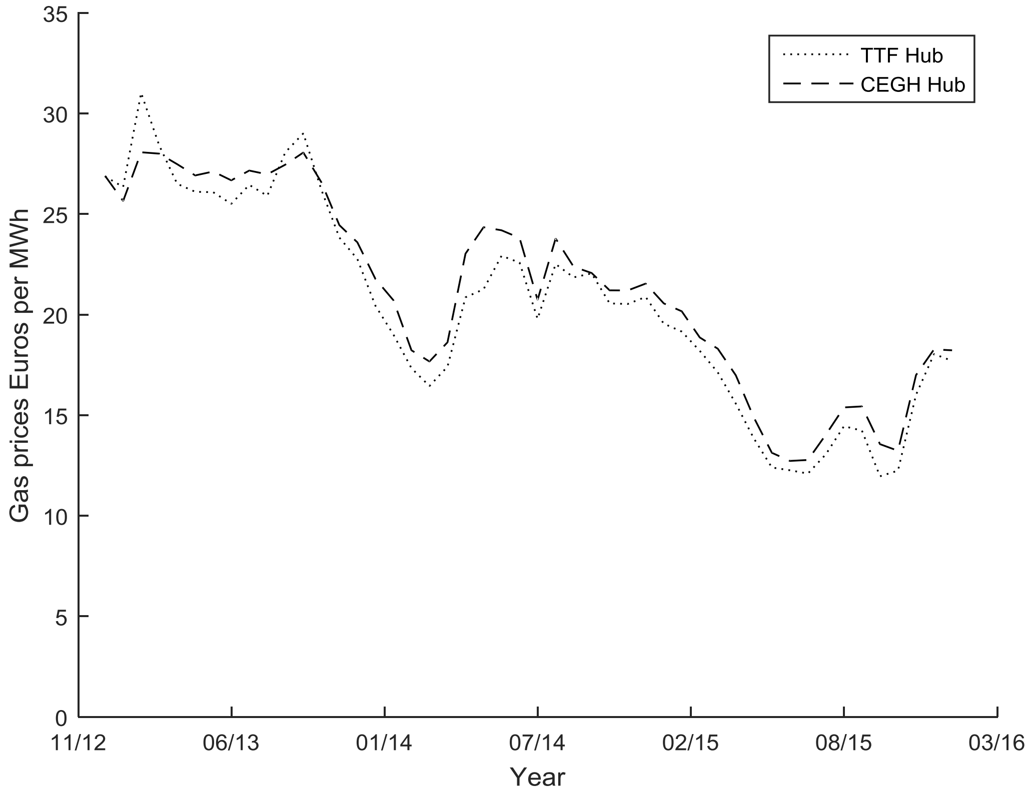

Figure 10. Monthly Gas Prices for the two Gas Hubs, Euros/MWh[26]

I start by calculating the market concentration, which is a function of the number of the gas suppliers and their respective shares of the total traded gas inside Austria and The Netherlands. In order to compute the concentration index, the Herfindahl-Hirschman index, HHI is used.

$$HHI_t = \sum\nolimits_{i=1}^{I} Q_{i,t}^2 \Big/ \left(\sum_{i=1}^{I} Q_{i,t}\right)^2 \qquad (1)$$

After computing the concentration in each market and at different time periods, two different approached will be used: parametric (Section 5.3.1) and non-parametric (Section 5.3.2).

5.3.1 Parametric approach

The parametric method takes into account the classical Nash-Cournot equilibrium test, with assumptions on cost and demand functions. For the sake of illustration, a very simple version of Cournot's model will be

[26] Available upon request for research students from Market and information services at the European Energy Exchange, marketdata@eex.com

considered, where two firms $\{i, j\}$ compete for gas quantities, and return to variations in the gas model in subsequent paragraphs.

Translated into modeling, each firm has a set of strategies and its space can be represented as $S_i = [s_{1,} s_{2,} \dots, s_n]$. In my case, s_i is a quantity choice, $q_i \geq 0$. The payoff $u_i(s_i, s_j)$ in a general two-player game can be written as:

$$u_i(q_i, q_j) = q_i * P(q_i + q_j) - C(q_i) \tag{2}$$

Where P is the price of gas, which is a function of total gas traded and C is the cost function.

According to Nash and as seen in Figure 11 the strategy pair (q_i^*, q_j^*) is a Nash equilibrium if, for each player i,

$$u_i(q_i^*, q_j^*) \geq u_i(q_i, q_j^*) \tag{3}$$

In the illustrative example, the equivalent statement is that each frim i solves the following optimization problem

$$\max_{q_i \in S_i} u_i(q_i, q_j^*) = \max_{q_i \in S_i} q_i * P(q_i + q_j) - C(q_i) \tag{4}$$

Firm j

Firm i

	q_j^{min}	q_2	q_3	q_j^*	.	.	q_j^{max}
q_i^{min}	(0,0)						$(0, U_j(0, q_j^{max}))$
q_2							
q_3							
q_i^*				Nash-equilibrium			
.							
.							
q_i^{max}	$(U_i(q_i^{max}, 0), 0)$						

Figure 11. Payoff matrix

$$(U_i(q_i^*, q_j^*) \geq U_i(q_i, q_j^*), U_j(q_i^*, q_j^*) \geq U_j(q_i^*, q_j))$$

Back to the parametric gas model developed in this study, the data is fitted to a classical linear Price function $P_t = a_t Q_t + b_t$, where $t \in \{2013, 2014, 2015, 2016\}$ represents the corresponding year. The latter means that each year has a different price function. The assumption of a simple linear regression model results in a low correlation coefficient. I acknowledge that there are many other explicative variables affecting the gas price and that it is more accurate to estimate the demand function using econometric models of high complexity level (i.e., vector autoregressive analysis using several supply and demand variables, machine learning, etc.). Due to the limited set of available data points, the fact that modeling the price function exceeds the scope of this chapter, and the fact that the use of the parametric method is only limited to comparison purposes, the assumption of a simple linear regression model is made for the demand function.

Furthermore, the optimal yearly quantities of each supplier are obtained by solving numerically the following set of equations:

$$Q_{i,t} \in argmax_{q\geq 0}\left\{q\left(\hat{a}_t^*\left(q+\sum_{j\neq i} Q_{j,t}\right)+\hat{b}_t^*\right)-C_{i,t}(q)\right\} \quad (5)$$

The estimated parameters $\hat{a}_t$ and $\hat{b}_t$ of each price equation are adjusted to obtain a price elasticity of demand equivalent to -0.2[27]. The adjusted parameters will be denoted by $\hat{a}_t^*$ and $\hat{b}_t^*$.

At each observation t, supplier i choses quantity q_i to maximize its profit given the output of other supplier, at its optimal choice, $Q_{j,t}$. The first-order condition for firm $i's$ optimization problem is compulsory; it yields to a classical optimization problem which is the context of a nonlinear programming. In addition to that, all the conditions of Krush-Kuhn-Tucker (KKT) are satisfied with a regular objective function to be optimized. Then, the simplest way to solve such kind of optimization problem is by applying the KKT principle.

$$Q_{i,t} * P_t'(Q_t) + P_t(Q_t) - C_i'(Q_{i,t}) = 0 \quad (6)$$

Where the first derivative of the inverse demand function is $P_t'(Q_t)$ and $C_i'(Q_{i,t})$ is the marginal cost of the gas. The final step of the optimization is to solve the system of nonlinear equations. The number of equations depends on the number of suppliers that are competing[28].

The production cost function of each supplier, in addition to the transportation costs, follow the form proposed by the following articles [71,114]. Equation 7, gives the capacity utilization marginal cost function $C_i'(Q_{i,t})$:

$$C_i'(Q_{i,t}) = \alpha_i + 2 * \beta_i * Q_{i,t} - \gamma_i \ln(1 - Q_{i,t}/Q_{max,i}) \quad (7)$$

$Q_{max,i}$ is the maximum production capacity of each supplier.

The parameters[29] α_i, β_i *and* γ_i are adjusted for inflation in order to fit for the purpose of the study, and are listed in Table 12. The production cost functions are convex and monotonically increasing [71,114,115].

Gas origin	**Gas yearly prices (USD per toe)**			
	α_i	β_i	γ_i	$Q_{max,i}, Annual\ proc$
Gas from Russia (Gazprom)	22	0	-41	100
Gas from Norway/UK	69	1	-18	80
Gas from the Netherlands (indigenous)	5	0	-22	70

Table 12. Cost assumptions parameters

[27] The price elasticity of demand is assumed to be -0.2 for both markets Austria and Netherlands. This is an assumption and different values that range between -0.1 to -0.8 are found in literature

[28] In the case of two competing suppliers, there will be two equations and two unknowns.

[29] For the data related to cost of production and the value of the parameters, please refer to the studies of [71,114] and [115], more specifically the section related to cost of production.

5.3.2 Non - parametric approach

A practical non-parametric method to test if a data set with convex cost functions belongs to Cournot, was developed by [73–76]. The authors developed a test that does not make any prior parametric assumptions about demand and cost curves, and thus allows for greater freedom in modeling.

$\delta_{i,t}$ is defined as an array of marginal cost of supplier i at period t. The first-order condition of firm $i's$ optimization problem previously defined in equation 6, now states that there is $\delta_{\mathrm{i},t} \geq 0$ contained in $C_i'(Q_{i,t})$ such that

$$Q_{i,t} * P_t'(Q_t) + P_t(Q_t) - \delta_{\mathrm{i},t} = 0 \tag{8}$$

The set of observations $\mathcal{C} = \left\{P_t, \left(Q_{i,t}\right)_{i\in\mathcal{I}}\right\}_{t\in\mathcal{T}}$ respects the Cournot equilibrium if both conditions mentioned below are met:

1) The observed price at each period should belong to the inverse demand function; this follows that the array $\{\delta_{i,t}\}_{(i,t)\in\mathcal{T}*\mathcal{I}}$ must obey the following condition

$$(P_t - \delta_{1,t})/Q_{1,t} = (P_t - \delta_{2,t})/Q_{2,t} = \cdots = (P_t - \delta_{I,t})/Q_{I,t} \geq 0 \; \forall t \in \mathcal{T} \tag{9}$$

2) At each period, firm i's quantity level $Q_{i,t}$ maximizes its profit subject to the quantity of the competing firms. This property can be stated as the following inequality

$$\left(\delta_{i,t'} - \delta_{i,t}\right)\left(Q_{i,t'} - Q_{i,t}\right) \geq 0 \; \forall t \neq t' \in \mathcal{T} \text{and} \; \forall i \in \mathcal{I}. \tag{10}$$

These conditions are derived from equation 6 as well.

Condition 1 compares the values of marginal costs of competing firms. This means that the marginal cost of the firm with the higher marginal cost will produce a quantity that is lower than the firm with the lower marginal costs.

Condition 2 allows us to compare the costs of different firms at different times. For instance, if firm i changes the produced quantity from $Q_{1,t}$ to $Q_{1,t'}$ the marginal cost at time t' must be lower than the marginal cost at time t. The same analysis for firm j leads to an arrangement of marginal costs for each firm and at each time in increasing order.

To make the problem of detection of Cournot's equilibrium, a numerical algorithm was developed. This algorithm is based on the results of the previous statements which indicate, in case of convergence, that the set of observations considered respects Cournot's equilibrium.[30] The algorithm starts with the highest possible marginal cost of firm i at $Q_{i,t}$ which is equal to the price P_t[31], and checks if conditions 1 and 2 are met.

This reasoning is in line with Cournot's Theorem which states that, as the number of firms in a market N, goes to infinity, the price P, converges to marginal cost.

$$\lim_{N\to\infty} P = marginal\ cost \tag{11}$$

This procedure is repeated in several iterations by changing the marginal cost each time until conditions 1 and 2 are fully met for a given cost value. If the algorithm does not converge, this means that the set of observations C does not respect a Cournot equilibrium. The algorithm is explained as follows:

- ***Step 1:*** *Consider the prices at time t (P_t) as a starting point of the upper bound of the marginal cost ($\delta_{i,t}^{ub}$) of the supplier i at time t. I.e. $MC_{i,t}^{ub} = P_t$.*
- ***Step 2:*** *Define for each supplier i and at time t the variable $\gamma_{i,t}$:*

$$\gamma_{i,t} = min\left\{\min_{\{t'\neq t:\, Q_{i,t'}>Q_{i,t}\}}\{\delta_{i,t'}^{ub}\}, \delta_{i,t}^{ub}\right\}$$

Note that if $\{t' \neq t: Q_{i,t'} > Q_{i,t}\} = \phi$ set $\gamma_{i,t}^{ub} = \delta_{i,t}^{ub}$.

This step ensures that Condition 1 is respected

- ***Step 3:*** *Define for each supplier i and at each time t the variables $\lambda_t$and $\gamma_{i,t}^{ub}$:*

$$\lambda_t = \max_j\left\{\frac{P_t - \gamma_{j,t}}{Q_{j,t}}\right\}$$

and

$$\gamma_{i,t}^{ub} = P_t - \lambda_t Q_{i,t}$$

This step ensures that Condition 2 is respected

- ***Step 4:***

  i. *If $\exists(i,t)$ such that $\gamma_{i,t}^{ub} < 0$, then the algorithm is stopped, and it is concluded that Cournot equilibrium conditions are not satisfied.*

  ii. *If $\forall(i,t)$, $\delta_{i,t}^{ub} = \gamma_{i,t}^{ub}$, then the algorithm is stopped, and it is concluded that Cournot equilibrium conditions are satisfied.*

[30] To get a better idea of the algorithm, derivation, proofs and lemmas, the readers are invited to go over the following papers [73–76]. However, the customized algorithm developed in this study is explained in this section.
[31] This is typical in a fully competitive market, where the price of any commodity (i.e. gas in this case) should be equal to its delivery cost.

iii. *Otherwise, return to Step 1 for a new iteration by letting* $\delta_{i,t}^{ub} = \gamma_{i,t}^{ub}$ *for all* (i, t).

- ***Finally**, if the algorithm does not stop at or before iteration T (number of periods), it is stopped by force, and it is concluded that Cournot equilibrium conditions are not satisfied.*

The Cournot acceptance rate is then calculated, which is the ratio of observations where conditions 1 and 2 are met to the total number of observations. The higher the ratio, means that the firms are competing under the umbrella of the a Cournot competition, each trying to behave strategically and willing to maximize its profit, taking into account of what it believes a strategic output of its rival

In a final step, and in order to check if changes in the concentration of suppliers affect the movement in gas prices in each market, the Herfindahl-Hirschman index is computed HHI_t, as well as the correlation coefficient with the prices P_t:

$$correlation\ (P_t, HHI_t) \tag{12}$$

5.4 Results and discussion

The average concentration indexes for all the data in both markets show signs of a low level of competition, and the values of both hubs indicate that the gas markets are concentrated to a certain extent. However, when comparing both values shown in Table 13, the CEGH scores 0.65 and is thus much larger than the TTF market, which sits at a value of 0.31. These numbers are a clear indication that the latter market is moderately less concentrated than the former.

Hub	Concentration index	Correlation coefficient	p-value
CEGH	0.65	0.21	0.1501
TTF	0.31	0.70	$3.2 * 10^{-8}$

Table 13. Correlation coefficient and significance

This indicates that the Netherlands leads the process of gas market liberalization, while other central hubs that are constrained by limited supply are lagging. Typically, as shown in Figure 8, the production and imports in a country like Austria are dominated by a large, state-controlled company that has limited competition from other suppliers, which are active in Germany, a factor that gives rise to oligopolistic market behavior.

In order to test the significance of the correlation coefficient describing the relation between prices and concentration indexes in both markets, the classical student test is performed. The results of the test are listed in Table 13.

The p-value in the Austrian market indicates that the correlation coefficient, which stands at a low value of 0.21, is not significant, which is the reason why there is no linear correlation between prices and the degree of market concentration.

By contrast, the correlation coefficient at the Dutch market is of higher value and significance, and stands at 0.7. The causation relation between the change in wholesale gas prices and the concentration of market shares in the Dutch market highlights the market power that oligopolistic traders have and can exercise at a given time by simply considering different strategic quantities. Additionally, it is worth mentioning that the short-term price dynamics are also affected by variables such as weather, storage, exchange rates, and supply disruptions.

Results suggest that there is no clear causality during the years of 2013-2016 in the Austrian market, in which the number of suppliers is limited. Each supplier's market share is large enough for even a modest change in strategic quantity by one large supplier to have a noticeable effect on the market shares or incomes of competent rivals.

Moving to the results of the parametric approach (See section 5.3.1) and using the yearly gas supply and prices, the results of the Nash Cournot equilibrium are explained below.

As previously mentioned, the Austrian market is supplied by three main parties. Based on the assumptions listed in the previous section, the optimal yearly quantities of each supplier are obtained by empirically solving equation 5 with the use of KKT conditions. The strategic quantities that give the maximum return for all three suppliers in the Austrian market in 2013 are computed and listed in Table 14.

2013	Outputs of each supplier – Austria	
	Computed quantities *GWh*	Actual values *GWh*
Russia (Gazprom)	17,104	36,531
Germany	16,715	6,728
Indigenous	1,264	1,264

Table 14. Strategic output of each supplier - Austria – 2013

As can be seen, the computed quantities differ from the actual values. According to the complementarity solution and based on the demand and cost assumptions, these values are the result of the strategies that resulted in the equilibrium solution of the different companies involved in the Austrian gas market.

It is important to note that the quantities are different when compared to the actual values. My model shows that in order to reach a Nash equilibrium for competing suppliers, more gas quantities have to be imported from Germany. The final strategic quantities for each supplier are then inserted into the inverse demand function computed using the regression technique in order to calculate the final wholesale price. These prices are shown in Table 15. The same analysis is performed for the Dutch market.

Hub	Gas yearly prices (USD per toe)			
	2013	2014	2015	2016
TTF	343	251	254	174
CEGH	340	297	271	193

Table 15. Parametric method - yearly gas prices results (in USD per Tons of Oil Equivalent)

Actual yearly average prices for the same years are shown in Table 16 below:

Hub	Gas yearly prices (USD per toe)			
	2013	2014	2015	2016
TTF	348	268	253	179
CEGH	348	285	264	190

Table 16. Actual values - yearly gas prices results (in USD per Tons of Oil Equivalent)

The results presented in Table 15 are comparable to the actual values in Table 16. The Nash equilibrium for all suppliers can lead to a price that is slightly lower than the actual prices in a certain year and higher in another. It is not easy to interpret such results. On the one hand, the resulting prices are similar to the actual prices, which indicates that the authors' model and assumptions are accurate to some extent. On the other hand, these prices are the result of strategic quantities that are optimized and that are not to be compared with actual values. The parametric method can indeed compute the optimal quantities of each supplier and subsequently the gas clearing prices. However, this requires assumptions on the market price, demand function, strategy of each supplier and the cost functions thereof. For this reason, the interpretation of the parametric results may be misleading and cannot be validated. To set these constraints free, it was decided to use a non-parametric approach that was previously explained in section 5.3.2.

The Non- parametric Cournot test results are shown in Table 17 and indicate the following:

Gas Hub	Cournot Acceptance rate (%)
TTF	50
CEGH	54

Table 17. Cournot acceptance rate

The behavior of large gas suppliers in both markets can be explained by a Cournot model, in which suppliers follow rational economics and try to maximize their payoffs. This is justified as more than 50% of the observations $\mathcal{C}$ respect the Cournot equilibrium, which means that the suppliers in that market are acting strategically to maximize their profits.

Interestingly both markets generate similar results, which mean that the suppliers in both markets follow the same behavior even though the Herfindahl-Hirschman index suggests the opposite.[32]

There should be a reason why the acceptance rate does not converge to higher values, and it could be that suppliers have other strategies in mind such as collusion, as in oil markets with OPEC, or other strategies that are not "pure" profit maximizers. This is in line with the findings of [37], who suggest that Gazprom is maximizing a 'utility function' relative to the choice of two variables/ objectives, one related to profit and another that is not related to profit. Others such as [40], suggest that certain European gas players are either

[32] As stated previously, the TTF is less concentrated than CEGH. In theory, it means that the behavior of the gas suppliers in the former market should show signs of competition.

not fully Cournot players, or that the new regulations and perhaps the old legacy contracts that are still ongoing prevent them from exerting full market power.

Although the CEGH is less dynamic due to its few suppliers, the game theory results for both markets (CEGH and TTF) are quite similar, which is a clear indication of market integration. This is in line with several other studies, namely [43,112]. In addition, the correlation of both hub prices is calculated and results in a value of 0.98; close to 1. The reasons for the lack of full and complete market integration are briefly summarized below:

Firstly, there may be storage problems that cause prices to increase, compared to adjacent markets, thus resulting in a decrease in correlation.[33] Secondly, there is a lack of transport capacity, which creates bottlenecks at times. Finally, the existence of contractual congestion and the concern about capacity hoarding are a main barrier to market integration.[34]

The high degree of market integration across Europe and the presence of strong liquid hubs in the northwest of the continent are bound to reduce the risk of market manipulation in the eastern part of Europe, where the gas hubs are less liquid and have a limited number of suppliers.

It is evident that due to the weak integration between gas markets and the lack of effective pricing mechanisms, suppliers in such markets can greatly influence the gas prices while the markets are isolated from trading opportunities arising in nearby countries. This is true for the European Baltic states, where Gazprom uses its dominant position to charge unfair prices [116]. [35]

[33] If the price differential between the two markets is less than the transportation fees, there is a risk of arbitrage, and the prices do not return to equilibrium.

[34] Eni, the Italian multinational oil and gas Company, has long-term contracts to use 85%-95% of Transitgas and TAG capacity and 67% of TENP capacity. Transitgas and TENP are gas pipelines that link Italy to Switzerland, Belgium, and the Netherlands on the one hand, and TAG links Italy and Austria on the other hand. Eni did not free up any unused capacity to either Transitgas or TAG. For this reason, the Italian antitrust regulator has opened an investigation for abuse of dominance position.

[35] The European Commission is contesting that Gazprom hindered the free flow of gas to the Baltic States.

Chapter 6

6. Northwestern European wholesale natural gas prices – Comparison of several parametric and non-parametric forecasting methods – Challenge 3

6.1 Nomenclature

$\mathrm{T_k}$	…	Average Temperature for a given k
k	…	Time t, equivalent to one week
$\mathrm{T_{ref}}$	…	Reference Temperature
i, j	…	Observations
y_i	…	Dependent/ explained variable
x_i	…	Independent/ explanatory variable
ε	…	Random disturbance, white noise process
p	…	Number of lags
$\mathrm{\beta_i}, \mathrm{\emptyset_i}$	…	Regression parameters
$\mathrm{C_i}$	…	Regression constants
$\mathbb{E}[\,]$	…	Expected value
$N(\mu, \sigma^2)$	…	Normal distribution with mean μ and variance σ^2
X	…	Real random variable
H_0	…	Null hypothesis
R^n	…	N-dimensional space
n	…	Positive integer
EC_t	…	Error correction vector
d	…	First difference
Δ	…	Double difference
α	…	Coefficient of the co-integration vector
θ		Coefficient of the Vector Error correction model

Table 18. Notation for the parametric methods applied in chapter 6

i	…	Number of input neurons
j	…	Number of hidden neurons
k	…	Number of output neurons
X_i	…	Input unit
x_i	…	Input vector $x = (x_1, \dots, x_i, \dots, x_n)$
Z_j	…	Hidden unit
z_j	…	Hidden unit vector $z = (z_1, \dots, z_i, \dots, z_n)$

Y_k	...	Output unit
y_k	...	Output vector $y = (y_1, \dots, y_i, \dots, y_n)$
W	...	Weight matrix: $W = \{w_{ij}\}$
E	...	Epoch (one forward pass and one backward pass of all the training examples)
ε	...	Very small number (close to zero)
V	...	Weight matrix: $V = \{v_{ij}\}$
v_{ij}	...	Weight on connection from unit X_i to unit Z_j
w_{jk}	...	Weight on connection from unit Z_j to unit Y_k
v_0	...	Initial value
z_in_j	...	Output of input unit vectors: $z\_in_j = v_0 + \sum_i x_i v_{ij}$
$f(z_{in_j})$	...	Activation function of hidden units Z_j
y_in_k	...	Output of hidden unit vectors: $y\_in_k = \sum_j z_j w_{jk}$
$f(y_{in_k})$	...	Activation function of output units Y_k
$w_{.j}$	...	Vector of weights: $w_{.j} = (w_{1j}, w_{2j}, \dots, w_{nj})^T$. This is the j^{th}column of the weight matrix
t_k	...	Training (or target) output vector $t = (t_1, \dots, t_j, \dots, t_m)$, in other words, it is the original price of gas
δ_k	...	Portion of error correction weight adjustment for w_{jk}that is due to an error at output unitY_k. It is the information about the error at unit Y_k
δ_j	...	Portion of error correction weight adjustment for v_{ij}that is due to an error at output hidden unit Z_j. It is the information about the error at unit Z_j
α	...	Learning rate
$\Delta w_{jk}, \Delta v_{ij}$	...	Weight correction terms
		Least square optimization
$\emptyset_j(X)$	...	Output of the input layer
X	...	Input vector
μ	...	Center of Radial Basis Function
δ_j^2	...	Spread of the Gaussian function
$\lVert \rVert$	...	Euclidian distance between input vector and μ
$y_k(X)$	...	Output layer
W	...	Weight matrix: $W = \{w_{kj}\}$
W_{k0}	...	Bias and initial value
D_i	...	Distance between the training sample and the point of prediction

Table 19. Notation for non-parametric methods applied in chapter 6

6.2 Introduction

Energy commodity price volatility is of great concern to oil and natural gas market participants, as well as policymakers. Being able to accurately forecast this volatility carries direct implications for trading**.** Facts have shown that natural gas, which is traded on the wholesale market, exhibits particularly large increases in price volatility.

The subject of concern in this study is Germany, where the pressure to liberalize the whole sale market in Europe has broken the link of oil and gas prices. Nowadays, the spot wholesale gas markets play an important role in the overall energy/ commodity transactions.

Germany is the largest western European gas market and has direct physical connection between several of its neighbors, most importantly the Austrian and Netherland gas markets. There should be a strong linkage between these markets. Many studies deal with the impact of European integration on gas market: Recently, [70] measured the degree of integration of gas markets based on the prices of eight European hubs, finding a significant level of convergence.

The increased competition on the wholesale market can have an effect of squeezing the profit margins of traditional suppliers and thus make the investment at high risk premium; this could lead to particularly large increases in price volatility. Economic theory argues that price changes are rooted in changes in supply and demand factors as pointed out by [44].

Rational economic interpretation listed in chapters 1 and 3, more specifically the ones related to challenge three, contribute to the understanding of what causes natural gas price volatility in European gas hubs. This study aims at using mathematical modeling, more specifically a multi-linear causal regression to check if the supply and demand variables help in forecasting short term natural gas spot day ahead prices.

Four methods will be used to compute the coefficients of a multivariate causal regression analysis: least square, maximum likelihood, machine learning gradient decent and least square optimization.

6.3 Material and methods

6.3.1 Study area and data used

The data set is composed of natural gas' weekly prices from September 2007 to December 2014 of the German gas hub NCG. Table 20 summarizes the definition of the variables used in this study.

Variable	Frequency	Description	Unit	Source
Coal	Weekly	Coal spot price for North-Western-Europe	Euro per ton	Global Coal[36]
Heating Degree Days (HDD)	Weekly	Deviation from historical heating degree days (1990-2014)	Degrees Kelvin	Degree Days[37]

[36] Available upon request for research students from the Global Coal Marketing & communication officer, frontoffice@globalcoal.com

[37] Available at http://www.degreedays.net/

Brent Oil	Weekly	Weekly Europe Brent Spot Price FOB (Dollars per Barrel)	Euro per barrel	Energy Information Administration (EIA)[38]
Exchange rates	Weekly	Exchange rates, EUR/USD	Unit less	Exchange rates[39]
Storage	Weekly	Storage utilization rates for Germany	Percentage points	Gas Infrastructure Europe[40]
Natural Gas, NCG	Weekly	Net Connect Germany (NCG) day ahead natural gas spot price	Euro per Megawatt hour	European Energy Exchange (EEX)[41]

Table 20. Data collection - Econometric models

To better illustrate and explain the process of data collection, the choice of the supply and demand variables used in this study will be explained below.

Research articles contain extensive literature reviews for natural gas spot prices under consideration of influencing variables derived from heating degree days such [117]. The nonlinear characteristic was observed long ago and used to define the Heating Degree Day (HDD)

$$\mathrm{HDD_k} = \max(0, \mathrm{T_{ref}} - \mathrm{T_k}), \tag{1}$$

Where $\mathrm{T_k}$ is the average temperature for the $\mathrm{k^{th}}$ week, and $\mathrm{T_{ref}}$ is the reference temperature, historically set to 288.5 Kelvin.

The focus will be on the deviations from the normal seasonal meteorological pattern as a determinant of gas prices. Thus, in a first step, the historical average seasonal series of heating degree days (HDD) will be considered. In a second step, the deviations of observed HDD and their historical averages will be calculated, in order to estimate the effects of unexpected temperature conditions on gas prices.

Spot in oil and gas markets is defined as day ahead, whereas in the coal markets, which are less liquid, spot is defined as cargos for delivery in the prompt plus 1, 2 & 3 months. The spot prices of Brent for oil prices and the spot prices for the DES ARA (Delivered Ex Ship at Amsterdam, Rotterdam or Antwerp) market, which is the main hub for coal delivered into Europe, will be used as the substitute for gas.

There is not a clear cut between German spot market ad storage behaviour: There is a lot of seasonal storage booking, a fragment of it can be used for short term variation. The issue of defining the appropriate market to study price/storage interlink is:

In Western Europe (let's say where market functions), we have to take a wider perspective than one country; we may use a storage in country A to balance the customers in country B.

In Eastern Europe, markets are more nationalistic but there is no robust price reference to observe the link between storage and price

[38] Available at http://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=pet&s=rbrte&f=d
[39] Available at http://www.exchange-rates.org/history/EUR/USD/T
[40] Available at https://transparency.gie.eu/
[41] Available upon request for research students from Market and information services, marketdata@eex.com

This study will focus on the German storage data only, and the utilization rates will be considered instead of absolute volumes to take into account for changes in the total storage capacity.

It is not easy to account for supply data, mainly because it is nearly impossible to expect the quantity to be delivered. This is why no variable related to supply interruption will be included and since there is a lack continuous data about LNG delivery, the supply variable will be omitted in the study.

Many other variables can potentially be included to reflect relevant political events, financial market conditions, extreme weather events like hurricanes, facility breakdowns causing supply shocks, and oil storage inventories. However, in this study, only include the variables explained above will be used.

6.3.2 Mathematical methods and data analyses

In order to characterize the dependent variable, which is the spot gas prices, some descriptive statistics were performed and it is clear that the mean and median have quit similar values; this means that this variable does not include outliers.

The descriptive statistics shown in Table 21, in addition to the frequency distribution of gas prices shown in Figure 12 lead us to conclude that the dependent variable is close to a normal distribution.

Variables	**N**	**Min**	**Max**	**Mean**	**Standard Deviation**	**Median**
Coal prices	378	57.7	219	98.3	30.1	87.82
Heating Degree Days (HDD)	378	266	298	277	6.15	275
Brent prices	378	35.4	142	95.2	21.9	103.1
Exchange rates	378	1.20	1.59	1.36	0.08	1.35
Storage	378	0.17	0.98	0.70	0.21	0.74
Natural gas prices	378	7.37	35.1	21.7	5.45	23.1

SD= Standard deviation, N= Number of samples

Table 21. Descriptive statistics

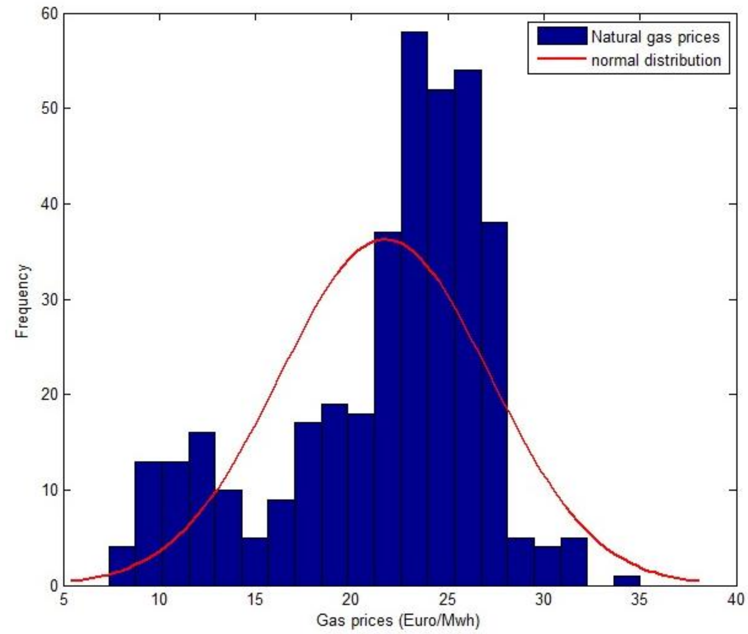

Figure 12. Frequency distribution of gas prices

Before analyzing the methods, the readers should be reminded of the aim of the study, which is to compute the coefficients of multivariate causal regression analysis and in a second step, to test the short-term prediction of natural gas prices for each method used. For this purpose, the time frame of the study is divided into two parts: September 24, 2007,to January 13, 2014, to compute the regression coefficients and the period of January 20, 2014, to December 15, 2014, to test the efficiency of the methods.

6.3.2.1 Least square method

The first method used in this study is the parametric multiple linear regression model. It is used to study the relationship between a dependent variable and one or more independent variables. It is assumed that each observation in a sample $(y_i, x_{i1}, x_{i2}, \ldots, x_{iK}), i = 1, \ldots, n$, is generated by an underlying process described by

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{iK}\beta_K + \varepsilon_i \quad (2)$$

Where y is the dependent or explained variable and $x_1, \ldots, x_k$ are the independent or explanatory variables.

The term ε is a random disturbance or a white noise process, it arises primarily when it is not possible to capture every influence on the dependent variable in a model, no matter how elaborated our model is. In the model the dependent variable is natural gas spot price and the independent variables are the supply and demand factors cited Table 21. The assumption of the stochastic process that has led to the observations of the data in hand are as follows:

- Linearity: $y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{iK}\beta_K + \varepsilon_i$. The model specifies a linear relationship between y and $(x_1, \ldots, x_K)$.

- Full rank: There is no exact linear relationship among any of the independent variables in the model. This assumption will be necessary for the estimation of the parameters of the model.
- Exogeneity of the independent variables: $E\ [\varepsilon_i\ |x_{j1}, x_{j2}\ , \ldots, x_{jK}]$ = 0. This means that the independent variables will not carry useful information while predicting ε_i.
- Homoscedasticity and no-autocorrelation: Each disturbance, ε_i has the same finite variance, σ^2 and is uncorrelated with every other disturbance, ε_j . $E\ [\varepsilon_i\ |\varepsilon_{i-1}] = 0$.
- Normal distribution: The disturbances or residuals are normally distributed. $\varepsilon|X{\sim}N(0, \sigma^2)$.

To test for normality and autocorrelation, the Jarque –Bera and the Durbin-Watson tests are used consecutively. The null Hypothesis for the model is:

$H_0 => Independent\ variable\ x\ is\ not\ useful\ in\ explaining\ the\ dependent\ variable\ Y, B_i = 0.$

6.3.2.2 Maximum likelihood methods

The second method is based on the principal of maximum likelihood using the method of vector autoregressive analysis, which is a parametric method.

A vector autoregressive is a system in which each variable is regressed on a constant and a certain p of its own lags as well as on p lags of each of the other variables under study.

Here is a large matrix notation of the VAR:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ \vdots \\ y_{k,t} \end{bmatrix} = \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_k \end{bmatrix} + \begin{pmatrix} \emptyset_{1,1}^{(1)} & \cdots & \emptyset_{1,k}^{(1)} \\ \vdots & \ddots & \vdots \\ \emptyset_{k,1}^{(1)} & \cdots & \emptyset_{k,k}^{(1)} \end{pmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ \vdots \\ y_{k,t-1} \end{bmatrix} + \cdots + \begin{pmatrix} \emptyset_{1,1}^{(p)} & \cdots & \emptyset_{1,k}^{(p)} \\ \vdots & \ddots & \vdots \\ \emptyset_{k,1}^{(p)} & \cdots & \emptyset_{k,k}^{(p)} \end{pmatrix} \begin{bmatrix} y_{1,t-p} \\ y_{2,t-p} \\ \vdots \\ y_{k,t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \vdots \\ \varepsilon_{k,t} \end{bmatrix} \tag{3}$$

Assumptions of the maximum likelihood method:

- It is essential to test the unit root and integration of our variables, therefore I shall apply the augmented Dickey–Fuller test (ADF): Indeed results show that some of the variables are non-stationary and to be more specific all commodity time series are non-stationary. In order to avoid inconsistency with model results as much as possible, the first difference for the following three main variables is included: coal prices, Brent prices and gas prices.

- The long-run relationships among the variables using the co-integration testing procedure suggested by [118] is tested, and statistical evidence of co-integration is found among some of the variables. The results of the test indicate that, there exists two co-integrated vectors, therefore the suitable estimation technique is a vector error correction model (VECM).

- Choosing the lags p: Based on Table 22, the p-value of the statistic test for lag count of two is less than 5% and more than 5% for lag values higher than 2. The results also indicate that the best (that is minimized) values of the respective information criteria, (Akaike criterion, Schwarz Bayesian criterion and Hannan-Quinn criterion) is reached at a lag count of two. Hence, by doing the likelihood-ratio test, the null hypothesis of the statistic test is accepted and a Vector Error Correction Model with a number of lags equal to two is considered.

Likelihood ratio test	*Chi-square (36)*
Lags: 2	0.0000
Lags: 3	0.1694
Lags: 5	0.2705

Table 22. Likelihood-ratio test results

- The advantage of the Vector Error Correction Model (VECM) from the linear regression analysis is that, exogenous and endogenous variables can be separated and this generates an accurate understanding of what drives the gas prices in the short and long term. Table 23, lists the exogenous and endogenous variables. One can argue that a harsh weather on a given day can indeed affect the gas prices, yet it is evident that a change of price will not necessarily affect the weather. Generally, the remaining variables correlate in such a way where the change in one will crop a result in the other.

Variables	*Description*	*Endogenous*	*Exogenous*
Coal prices	First difference	✓	
Heating Degree Day (HDD)	Heating degree days		✓
Brent prices	First difference	✓	
Exchange rates	Euro-US, exchange rates	✓	
Storage	Storage utilization rates	✓	
Natural gas prices	First difference	✓	

Table 23. List of exogenous and endogenous variables

It is worth mentioning that, in order to get to finalize the choice on which variable to include as exogenous and which one to designate as endogenous, some statistical trial and error work was done on “Gretl” software and it can be summarized as computing several VECM model combinations.

- To test for normality the Doornik–Hansen test was used. Therefore, the decision concerning the normality of residuals based on the p value can be made.

6.3.2.3 Gradient decent method

Before using the gradient decent method, the statistical procedure called principal component analysis (PCA) will be sued on the dataset. The main objective is to simplify the data structure by reducing the dimension of the data. The original variables are rearranged into several new uncorrelated comprehensive components (or factors) without losing significant information. Several studies have used PCA in multivariate causal analysis, such as [119] and [120].

6.3.2.3.1 Principal component analysis

Before using the gradient descent method, the data will be analyzed using a statistical procedure called principal component analysis (PCA). The main objective is to shorten the data set without losing significant information. The data are composed of five independent variables, each containing 378 observations, so one can think about a cloud of points in an n-dimensional space R^5, and PCA is employed to find and study the relationships among the data set of 1890 observations. The correlation matrix is shown in Table 24.

Variables	Coal prices	Heating Degree Days	Brent prices	Exchange rates	Storage	Natural gas prices
Coal prices	1	-0.082	0.466	0.597	0.217	0.402
Heating Degree Days (HDD)	-0.082	1	0.198	-0.218	-0.091	0.249
Brent prices	0.466	0.198	1	0.214	0.094	0.572
Exchange rates	0.597	-0.218	0.214	1	0.136	-0.129
Storage	0.217	-0.091	0.094	0.136	1	0.028
Natural gas prices	0.402	0.249	0.572	-0.129	0.028	1

Table 24. Correlation matrix

The PCA results of five supply and demand variables are shown in Table 25 and Figure 13. Four major principle components (F1 to F4) affecting the natural gas prices are identified; combined they explain 93% of the original data variance. The data shown in bold indicate higher loading and contribution to the corresponding components i.e. the high positive correlation between variable and component. Component, F1, which explains 39% of the total variance, have high loadings on coal, exchange rates, and Brent prices.

Variables	F1	F2	F3	F4
Coal prices	**0.635**	0.055	0.114	-0.078
Heating Degree Days (HDD)	-0.117	**0.787**	-0.148	-0.587
Brent prices	**0.438**	0.539	0.008	0.632
Exchange rates	**0.560**	-0.227	0.297	-0.495
Storage	0.278	-0.185	**-0.936**	**-0.068**

Table 25. Rotated factor loadings of principal components on variables

In the bi-plot, vectors represent the adjusted eigenvectors, and the points represent the principal component scores; that is, the representation of space vectors that point in the same direction. This means that the variables have the same response profiles.

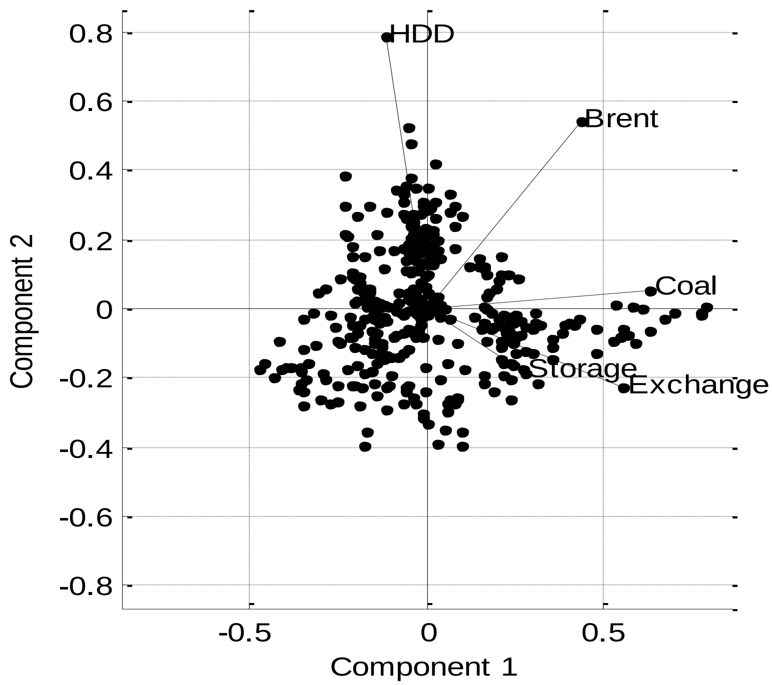

Figure 13. Bi-plot of the principal components factors (F1 and F2)

The length of the lines approximates the magnitude of the variances of each variable. The cosine of the angle between the lines represents the correlation between the variables; this means that the smaller the angle, the more the variables are correlated. Since the aim is to reduce the dimension with the variable that has a similar interpretation, consequently, the coal vector was omitted from the study.

Now that the dimension of the data is reduced, the NN gradient descent method can be run. This method is the machine learning nonlinear and non-parametric approach chosen in this study.

6.3.2.3.2 Gamma test

The usage of the Principal Component Analysis was useful in reducing the dimension of the data set and I succeeded to shorten our data space to 4 independent variables and one dependent variable, each consisting of 378 observations.

Artificial Neural Network is a non-parametric method that mostly use noon-linear activation functions in order predict and forecast commodity prices, and there is one vital question that remain to be answered. How many observations/ data points are needed to calibrate and forecast with accuracy?

Several researchers worked on developing mathematical models that can search for relevant data points in a big data set. The Gamma test is one example and an elaborated proof was completed by [121], and [122]. This method would allow users to pick specific observations of a time series and in order to achieve the best performance target. Once completed, the test, shall help the modelers overtake the over-parameterization and overtraining issue that is normally associated with the non-parametric methods, mainly the Neural Networks. In addition it will give us an estimate of how smooth the forecast model can be [123].

Considering a set of input-output data

$$\{x_1(i), \dots, x_4(i), y_i\} = \{(x_i, y_i) | 1 \leq i \leq M\} \tag{4}$$

Where the vector $X = (x_1, \dots, x_4) \in R^4$as the input in a closed bounded set $C \subset R^m$, and $y \in R$ as the output. M represents the number of observed series for underlying phenomenon. The common question is: How much is the output determined by the input?

The relationship between the four input vectors $x = (x_1, \dots, x_4)$ and the main output y natural gas prices vector can be decomposed into two distinct components as:

$$y = f(x_1, \dots, x_m) + r \tag{5}$$

Where f is a smooth function of the independent variables and r is the noise. It is assumed that the distribution of r has a mean of zero and a variance $Var\ (r)$. The gamma test finds an estimate for $Var\ (r)$, i.e. that part of the model's output variance which cannot be accounted for by the smooth model.

Although f is unknown, the Gamma test is still able to provide, under some reasonable, an estimate for $Var\ (r)$, "Gamma Value". The estimate of the component of the variance of the output not accounted for by a smooth model is alternatively called the $Gamma\ statistic$ and is denoted by α.

This specific target is called Gamma statistic, and it should give us the best mean-squared error of the predictions of $y^*(i)$ compared with the original output data $y(i)$.

$$MSError = \frac{1}{M}\sum_{i=1}^{M}(y^*(i) - y(i))^2 \tag{6}$$

Accordingly, the Gamma statistic with the lowest expected $MSError$[42] will be chosen and thus the time series data and number of observations will be reduced.

The Gamma test calculates the mean-squared pth nearest neighbor distances $\delta(p)(1 \leq p \leq p_{max})$ and the matching $\gamma(p)$ as defined next. Although, the Gamma test is an unknown function of f, it can directly estimate $Var\ (r)$ from data. So given data samples $\big(x(i), y(i)\big)$, where $(x(i) = (x_1(i), \dots\ x_m(i))$, $1 \leq i \leq M$, and allowing $N[i, p]$ be the list of (equidistant) pth nearest neighbors to $x(i)$. We have:

$$\delta_M(p) = \frac{1}{M}\sum_{i=1}^{M}\left|x_{N[i,p]} - x_i\right|^2 \tag{6}$$

Where $|.|$denotes Euclidean distance and $\gamma(p)$, is:

[42] For more on the Gamma test and its application, please refer to [137].

$$\gamma_M(p) = \frac{1}{2M}\sum_{i=1}^{M}\left(y_{N[I,P]} - y_i\right)^2 \ (1 \le p \le p_{max}) \tag{7}$$

Then, the fitted regression lines passes through points $\left(\delta_M(p), \gamma_M(p)\right)(1 \le p \le p_{max})$ points, like:

$$\gamma = A\delta + \alpha \tag{8}$$

One can standardize the result by considering another term V_{ratio}, which gives a scale invariant noise estimated between 0 and 1, a kind of the 'goodness of fit' of the data with respect to the class of smooth functions with bounded derivatives. V_{ratio}, otherwise known as Gamma statis, is defined as

$$V_{ratio} = \frac{\alpha}{\delta^2(y)} \tag{9}$$

Where $\delta^2(y)$ =variance of output y, which allows a judgment to be formed independent of the output range as to how well the output can be modeled by a smooth function. A value of V_{ratio}close to 0 indicates that there is a high degree of predictability of the given output y. In fact,

$$1 - |V_{ratio}| = 1 - \frac{|\alpha|}{Var\ (y)} \tag{10}$$

is closely analogous to the conventional r^2statistic which estimates the extent to which the data fits a linear model, except that here V_{ratio}estimates the extent to which the data fits a smooth nonlinear model.

The M-test, otherwise known as the standard error, is a way to gauge whether the Gamma statistic estimates $Var(r)$ reliably. It performs by computing the Gamma statistic for a given subset of the available data. Whereby at each successive calculation of the Gamma statistic I increase M by small step, until I have either used all the data or the statistic has converged enough towards a fixed value.

The standard error of regression line is calculated as follows:

$$SE(\alpha) = \sqrt{\frac{1}{n-2}\sum_{i=1}^{p_{max}}(\alpha(i) - \bar{\alpha})^2} \tag{11}$$

Where, identifier i is the ith Gamma regression point value and $\bar{\alpha}$ is the mean. If the standard error value is close to zero, I have more confidence in the value of the Gamma statistic as an estimate for the noise variable on the given output. In our case the standard error is never 0.

In an attempt to reduce the training time for neural networks and improve the resulting model by eliminating irrelevant inputs into our model, the Gamma test is applied to the remaining 4 independent input vectors.

As can be seen in Figure 14, the quantity of data points required to construct a reliable model is determined using the V_{ratio} (Gamma static) and the M-Test (Standard error).

As can be seen the Gamma static is increasing and nowhere close to 0, which indicates that there is a high degree of unpredictability of the given output y. Additionally the standard error does not converge to zero, which gives less confidence in the value of the Gamma static. This simply means that the data set for each of the four variables are not enough to generate an accurate forecast according to the Gamma test. I used the latter model to reduce the data sets, but on the contrary the results are showing that I need more data in that case.

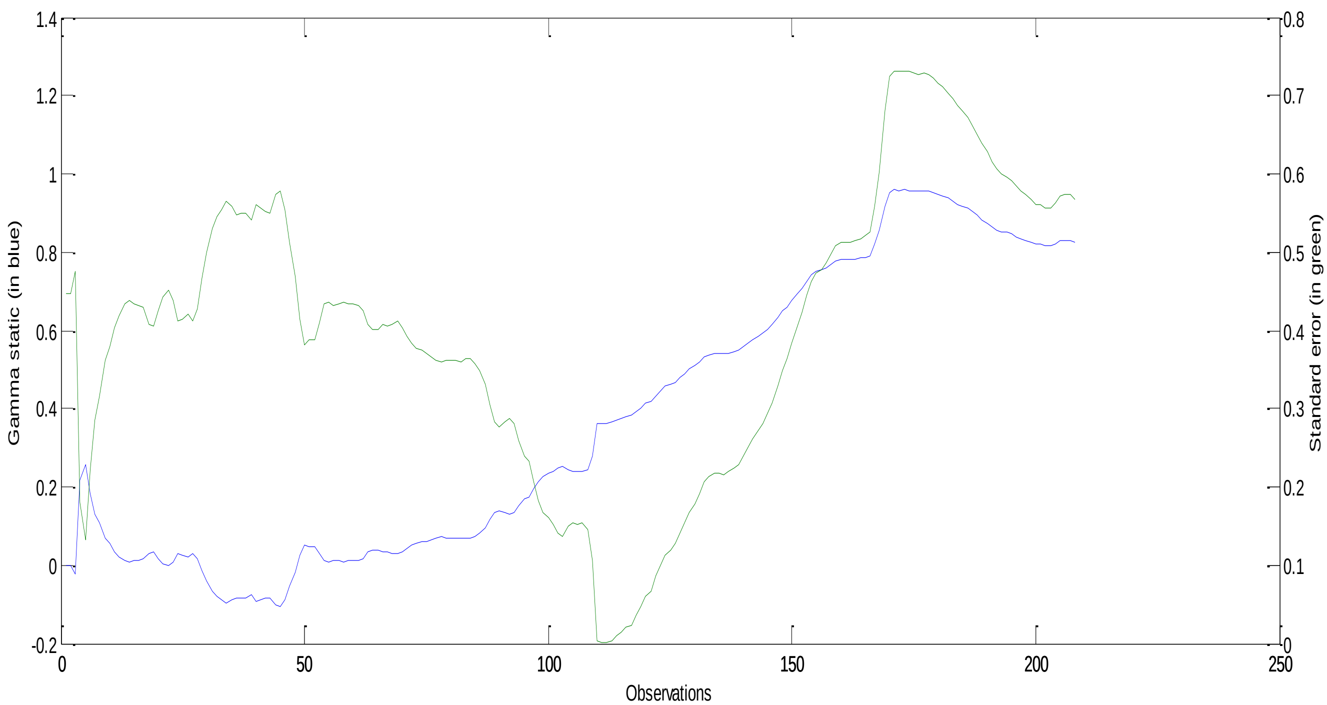

Figure 14. Observations vs. Gamma & Standard Error

As much as the results of the Gamma test can be important in any model, however the lack of solid results suggest that its application can be misleading at times, especially for the causal multivariate causal regression analysis. Univariate time series models that refer to time series that consists of single (scalar) observations, might work better for Gamma test, where the determination of which lagged variables irrelevant in predicting a certain output.

The test will therefore explore different combination of lagged variables as inputs, instead of completely different observation sets and variables as is the case the multivariate causal regression analysis.

A neural net consists of a large number of simple processing elements called neurons. Each neuron is connected to other neurons by means of directed communication links, each with an associated weight. The weights represent information being used by the net to solve the problem. Typically, a neuron sends its activation as a signal to several other neurons.

This neural network is characterized by: (i) its pattern of connections between neurons called its architecture, (ii) its method of determining the weights on the connections called its training or learning algorithm and (iii) its activation function (usually nonlinear) applied to the net input (sum of weighted input signal) of each neuron to determine its output signal

The arrangement of neurons into layers and the connection patterns within and between layers is called the net architecture (shown in Figure 15).

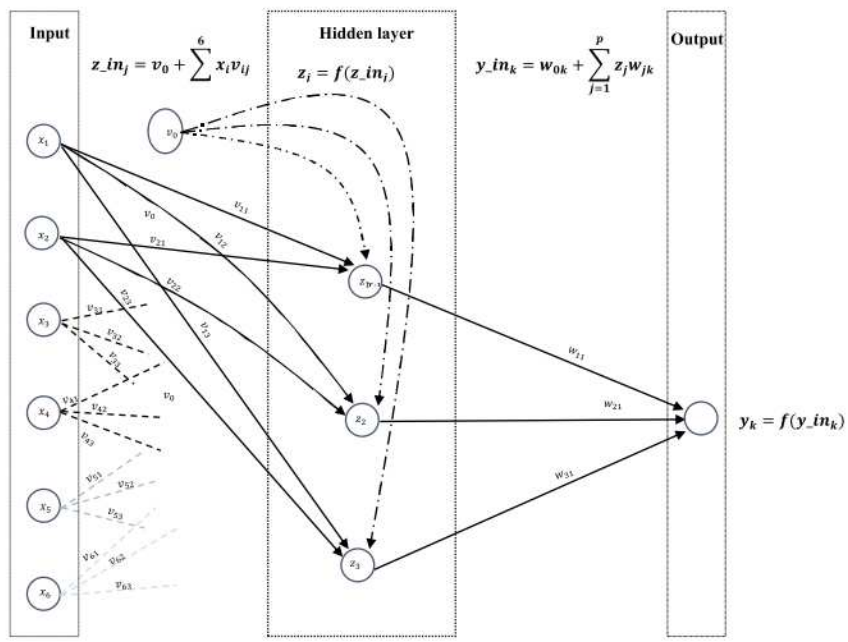

Figure 15. Back propagation neural network structure illustrative drawing

Assumptions of the gradient decent method:

- Input vectors contain all the data related to the explanatory variables data used to predict the gas price. Therefore the input layer contains 4 neurons.
- In this study, the two-layer feed-forward network will be used; the 1st layer, “hidden layer”, contains 4 neurons whereas the 2nd layer, “output layer” is formed of one neuron. The Gas price data forms the output vector.
- The authors of this study tried to choose graphically the number of hidden layers; unfortunately, the data are not showing any clusters that are separable by hyper-plane; thus, one hidden layer will be considered.
- Because there is no significance for the constant and bias in the Multi linear regression it is assumed that the bias is equal to zero (so v_0 & w_{0k} are both set to zero)
- The nets used in this study are feed forward nets (nets in which the signals flow from the input units to the output units, in a forward direction)

  The gradient decent back propagation to train multilayer nets will be used. The back propagation computation is derived using the chain rule of calculus that is found in [124]

  Therefore, the next step is to minimize the error term, and to do so, the network will be trained using the gradient descent back propagation algorithm, so it can learn the underlying relationship between inputs vectors and the target vector (in other words adjusting the weights). The application divides data into 3 sets:

- 70% are used for training
- 15% are used to validate that the network is generalizing and to stop training before over fitting occurs
- The last 15% are used as a completely independent test of network generalization.

- As we are dealing with a nonlinear set of data, three different nonlinear activation functions will be used to introduce nonlinearity:

    - The Log-sigmoid function

$$logsig\ (n) = \frac{1}{1 + \mathrm{e}^{-\mathrm{n}}} \tag{12}$$

    - The Tan-sigmoid function

$$tansig\ (n)\ = \frac{2}{(1 + e^{-2n}) - 1} \tag{13}$$

    - The radial basis function

$$radbas\ (n) = \ e^{-(n^2)} \tag{14}$$

- The "hidden layer" $f(z\_in_1)$, has three different transfer functions whereas the "output layer" has a $purelin$ transfer function$g(f(z))$. The sigmoid/tansig/radbas function has been used for the following reasons:

  These activation functions have the added benefit of having simple derivatives which make learning the weights of a neural network easier. Although theoretically any differentiable function can be used as an activation function, the above mentioned functions are the most commonly used.

  In addition to that, the chosen functions are the most popular nonlinear functions used to explain such type of data in the practical applications.

- A common procedure is to initialize the weights to random values between -0.5 and 0.5. The choice of initial weights will influence whether the net reaches a global (or only a local) minimum of the error and, if so, how quickly it converges. The values for the initial weights must not be too large, or the initial input signals to each hidden or output unit will likely fall in the region where the derivative of the activation functions have a very small value. On the other hand, if the initial weights are too small, the net input to a hidden or output layer unit will be close to zero, which causes also extremely slow learning.

- The stopping condition is when at an epoch E, the norm of gradient of the error, is less than a given small positive value ε. When the validation error increases for a specified number of iterations, training is stopped, and the weights and biases at the minimum of the validation error are returned. In our case the training stops when the validation error increases for six iterations.

- The output of the input layer is determined by the following function, where v_{ij} is the vector of weights

$$z\_in_j = v_0 + \sum_{1}^{5} x_i v_{ij} \tag{15}$$

- The activation function will compute the following equation (in this case I will show the example of $logsig$ function)

$$z_j = f(z\_in_j) \tag{16}$$

$$f\left(z_{in_1}\right) = \frac{1}{1 + e^{-(v_0 + x_1 v_{11} + x_2 v_{21} + x_3 v_{31} + x_4 v_{41} + x_5 v_{51})}} \tag{17}$$

- The output of the network is determined by the following function

$$y_{\_in_k} = \sum_{1}^{3} w_k * y\_in_k \tag{18}$$

$$f\left(y_{in_k}\right) = \frac{w_{11}}{1 + e^{-(v_0 + x_1 v_{11} + x_2 v_{21} + x_3 v_{31} + x_4 v_{41} + x_5 v_{51})}} + \cdots + \frac{w_{31}}{1 + e^{-(v_0 + x_1 v_{17} + x_2 v_{27} + x_3 v_{37} + x_4 v_{47} + x_5 v_{57})}} \tag{19}$$

- As mentioned earlier, training a network by gradient back propagation involves the following stages:

  - *During feed forward*, each input unit (X_i) receives an input signal (in our case the input signal for each variable is equivalent to 70% of the initial number of samples listed in Table 21, and broadcasts this signal to each of the hidden units $Z_1, \dots, Z_n$. Each hidden unit then computes its activation and sends its signal (z_j) to each output unit. The output unit (Y_k) computes its activation (y_k) to form the response of the net for the given input pattern.
  - *During training*, each the output unit compares its computed activation y_k with its target value t_k to determine the associated error for that pattern with that unit. Based on that error, the factor δ_k (k=1 in my case) is computed. δ_k is used to distribute the error at output unit Y_k back to all units in the previous layer (the hidden units that are connected to Y_k). *The error information term* is defined as follows: $\delta_k = (t_k - y_k) f'(y\_in_k)$, (where the last term is the derivative of the output function) and is used later to update the weights between the output and the hidden layer.

- In a similar manner, the factor δ_j(j=1,…, p) is computed for each hidden unit Z_j. It is not necessary to propagate the error back to the input layer, but δ_j is used to update the weights between the hidden layer and the input layer. This is done by multiplying the derivative of its activation function to calculate its error information term: $\delta_j = \delta\_in_j f'(z\_in_j)$, where δ_j is the portion of error correction weight adjustment for v_{ij}.

- *The weight correction terms* are used to update w_{jk} and v_{ij}, $\Delta w_{jk} = \alpha\delta_k z_j$ *and* $\Delta v_{ij} = \alpha\delta_j x_i$, where α is the learning rate that is used to control the amount of weight adjustment at each step of training.

- And the final step is to update weights and biases, the output unit (Y_k) updates its bias and weights (j = 0,.., m): $w_{jk}(new) = w_{jk}(old) + \Delta w_{jk}$, and ach hidden unit ($Z_j$, j = 1,…, m) updates its bias and weights (i = 0,…,n): $v_{ij}(new) = v_{ij}(old) + \Delta v_{ij}$

It is also worth mentioning that each time a feed forward is initialized, the neural network parameters are different and might produce different solutions due to different initial weight, bias values, and different divisions of data into training, validation, and test sets. As a result, different neural networks trained on the same problem can give different outputs for the same input.

6.3.2.4 Least square optimization methods

The final method of coefficient estimation is the least square optimization and for that purpose I will use radial basis networks is chosen. The weights of this network are optimized using least mean square algorithm. Radial Basis Networks (RBN) is based on supervised learning, it can be trained in one stage rather than using the iterative process as in gradient decent and is also good at modeling nonlinear data

Assumptions of the least square optimization method:

- Input vectors contain all the data related to the explanatory variables data used to predict the gas price Therefore there is one input layer containing 4 neurons
- It has the same type of architecture in the sense of perceptron and connections
- The output of the input layer is characterized by the following function $\emptyset_j$, it is a nonlinear function of unit j, which is typically Gaussian of the form

$$\emptyset_j(X) = exp\left(-\frac{\| X - \mu \|^2}{2\sigma_j^2}\right) \tag{20}$$

Where X and μ are the input and the center of radial basis function (RBF) unit respectively. σ_j^2 is the spread of the Gaussian basis function and ‖‖ is the Euclidian distance between the input vector and the chosen centers of RBF. The centers can be chosen randomly using clustering algorithms. In this study, centers were randomly selected from the data set.

- The output layer of the network is derived from this equation,

$$y_k(X) = \sum_{j=1}^{M} W_{kj} \emptyset_j(X) + W_{k0} \tag{21}$$

- The weights are optimized using the least mean square algorithm once the centers of RBF units are determined. This is done in order to fit the network outputs to the given inputs.

- Three different models will be used in the study.

  - $The\ newrb\ model$: It iteratively creates a radial basis network one neuron at a time. Neurons are added to the network until the sum-squared error falls below an error goal (The mean squared error goal is set to 0) and a maximum number of neurons has been reached (The maximum number of neurons is set to 100)

  - $The\ newrbe\ model$: It creates as many $radbas$ neurons as there are input vectors. In other words, this function can produce a network with zero error on training vectors in which each $radbas$ neuron acts as a detector for a different input vector

  - $The\ grnn\ model$: The generalized regression neural network (GRNN) is often used for function approximation. The probability density function used is the normal distribution. Each training sample, is used as the mean of a Normal Distribution.

$$(X) = \frac{\sum_{i=1}^{n} Y_i e^{(-\frac{D_i^2}{2\sigma^2})}}{\sum_{i=1}^{n} e^{(-\frac{D_i^2}{2\sigma^2})}} \tag{22}$$

The distance, $D_{i,}$ between the training sample and the point of prediction is used as a measure of how well each training sample can represent the position of prediction, X. For $D_i = 0$, the exponential becomes one and the point of evaluation is represented best by the training sample.

- The same spread for all the models will used and is equal to five.

6.4 Results and discussion

6.4.1 Results of OLS method

It is observed from Table 26, that the p-value for the Exchange rates and the storage exceed 5%. Therefore these variables are not significant and they shall be eliminated from the model. In addition, the results show that the model has an R^2 of $59.3\ \%$ which indicates that the model explains fairly all the variability of the response data around its mean, and that the model fits the spot gas prices.

Variable	p-value
Intercept	0.001
Heating Degree Days (HDD)	2e-16
Brent prices	2e-16
Exchange rates	0.223
Storage	0.012
R-squared	0.593
Durbin-Watson stat	0.168
Jarque-Bera	9E-06

Table 26. Least square results

However, when testing for autocorrelation, the residual plot and the Durbin Watson test indicate that each disturbance ε_i is positively correlated with every other disturbance. In addition, the p-value of the Jarque-Bera Normality Test of residuals exceeds 5%, so the normality assumption is true.

From an economic point of view, the multi linear regression shows that, with the exception of the exchange rates and to a lesser extent storage, the other variables chosen in the assumptions have meaningful addition to the model because changes in the variables value are related to changes in the gas prices.

The higher the R^2, the better the linear model fits the data. However, in the first application, the Multi Linear regression model cannot be considered as the best fit for the simple fact that the assumption of normality and non-autocorrelation of residuals are not valid.

The second method is based on the principal of maximum likelihood using the method of vector autoregressive analysis

6.4.2 Results of the maximum likelihood method

The results show that the gas price is affected mainly by all variables included in the assumptions but for different periods. Also, by considering the VECM, results show that gas prices are directly affected by the lags of all other variables.

Table 27 and Table 28 summarize the parameter restrictions on the lagged relationships and the equation of regression for other variables in the VECM. On the basis of these results, it is noted that storage depend on its own lag values and on the weather (HDD); gas prices depend on lags of mainly all other supply and demand variables, although with smaller significance for storage and exchange rates; and on both error correction vectors (having in mind that each error correction vector is a linear combination of several variables that explain the co-integrating relationships).

Variable	p-value
Intercept	0.0006
Heating Degree Days (HDD)	0.0670
Brent prices	0.0164
Exchange rates	0.0586
Storage	0.0590
Gas prices	1.33e-05
EC1	4.1e-0.26
EC2	2.3e-0.41
R-squared	0.518
Durbin-Watson stat	2.017

Table 27. Maximum likelihood results

In the same time, the results show that the following variables also affect natural gas prices: coal prices, exchange rates, and its own lags.

The VECM model now can be described with the following mathematical annotation:

$$\Delta gas_{t=} constant + \beta_1 * HDD_t + \beta_2 * \Delta Storage_{t-1} + \beta_3 * \Delta Exchange_{t-1} + \beta_4 * \Delta Brent_{t-1} + \beta_5 * \Delta Gas_{t-1} + EC1_{\mathrm{t-1}} * \theta(1) + EC2_{\mathrm{t-2}} * \theta(2) * + dGas_{t-1} \quad (23)$$

where Δ denotes the double difference and d designates the first difference**,** β is the coefficients computed by the maximum likelihood method of the VECM, α is the coefficient of the co-integrating vector, and where the error correction vector $EC1_t$ is the co-integrating vector and θ is its computed coefficient and defined as follows:

$$EC_t = constant + \alpha_1 * HDD_t + \alpha_2 * dStorage_t + \alpha_3 * dExchange_t + \alpha_4 * dBrent_t + \alpha_5 * dGas_t \quad (24)$$

Because natural gas is affected by commodity prices and in turn, it also affects the commodity prices and exchange rates, this means that the model captures the economics of energy commodities.

The maximum likelihood-VECM method helps to understand the effects of various fundamental variables on gas prices, and the main conclusion that can be extracted from the method is that, unlike the second assumption in the multi linear regression, there is a clear relationship among the variables of the model.

	Exchange rates	HDD	Storage	Brent	gas	EC1	EC2
HDD							
Exchange rates	✓				✓	✓	✓
Storage		✓	✓				
Brent						✓	
Gas		✓		✓	✓	✓	✓

Table 28. Parameter restrictions on the lagged relationships using the VECM model for time (t-1) at a confidence level of 5 %

An additional critical test for the stability of the VECM is the unit root test. The results show that all the inverse roots are contained within the unit circle; this implies that the model is stationary over time.

Both methods used so far require assumptions about the relationship between the independent variables used to produce the gas prices, as well as the probability distribution of residual errors that has to be Gaussian. However, the results show that the VECM normality test failed as the p-value indicates that the normality assumption is rejected; also, at least one of the inverses of roots is almost equal to one. It is expected that the normality test failed because of an outlier in the distribution of residuals, and in the normal case, I can proceed using this model.

6.4.3 Results of the Gradient decent method

Figure 16 shows the plots of forecast of the three models ($logsig, Radbas\ and\ Tansig$); this is the result of the test period composed of 48 values. The predictive power of the gradient decent method and its relevant methods show that the forecasts are acceptable during the test period.

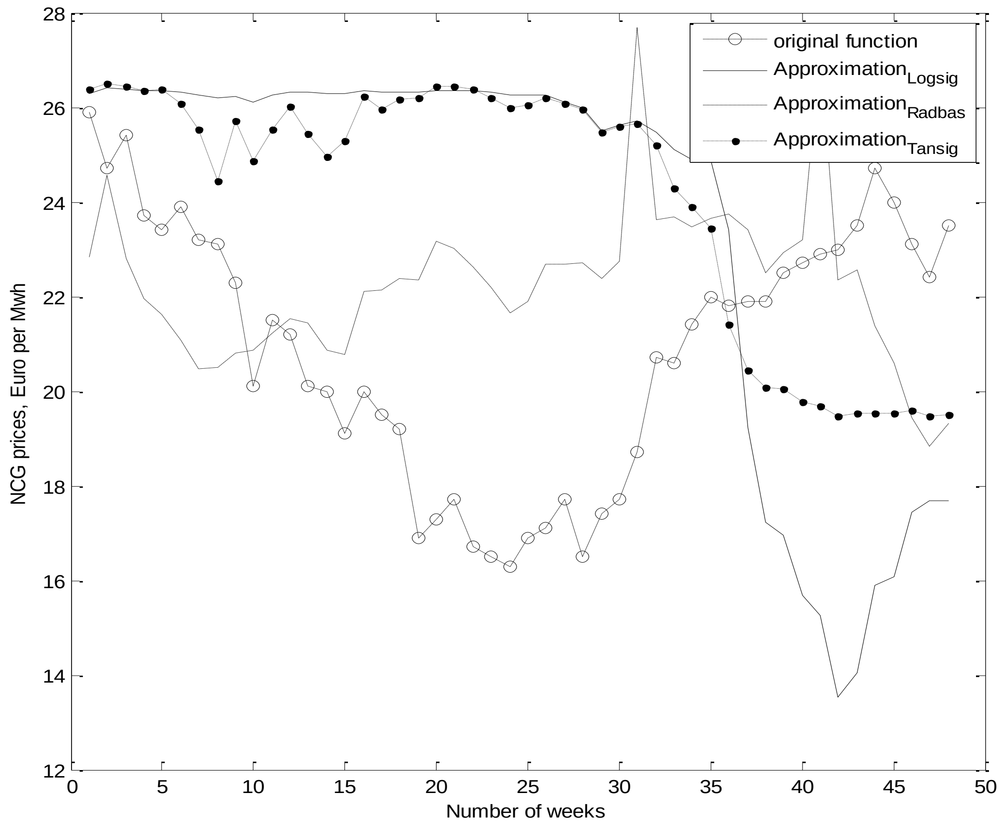

Figure 16. Plots of gas price forecasts using the Neural Network Gradient Decent method

6.4.4 Results of the least square optimization method

Figure 17 shows the plot of the function approximation and forecast of the $newrb$ model, which is one of the three models used in the least squares method. As can be seen, the RBN is accurate in approximating the function; however, it does not show the same accuracy during the test period.

Unlike the two previous forecasting methods that are parametric and constrained by several assumptions on the data, the gradient descent and least squares methods are "constraint-free". Therefore, there is no need for additional tests (i.e., normality test for residuals and autocorrelation). Also, machine learning forecasting methods perform nonlinear statistical modeling, which helps fitting complex relations in a given data set, and this is asserted in the results shown in Figure 17.

The results of all methods will be compared in the following section

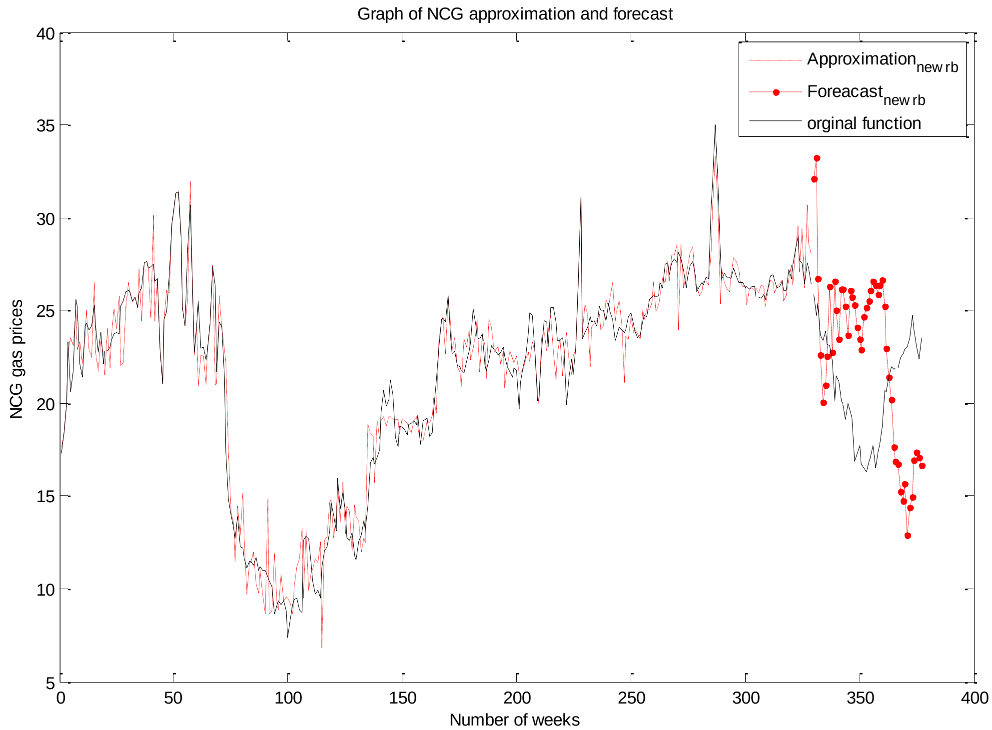

Figure 17. Plots of gas price forecasts using least square optimization method

6.4.5 Comparison analyses of all models

So far, eight different models were used to compute the coefficients of a multivariate causal regression analysis. In a final step, the mean absolute percentage error (MAPE) is solved to validate and compare the models.

$$Mean\ Absolute\ Percentage\ Error = \frac{1}{n}\sum_{i=1}^{n}\frac{Target\ price_{\ i} - Forecasted\ price_{\ i}}{Target\ price_{\ i}} * 100 \qquad (25)$$

The results are displayed in Table 29; the mean absolute error is computed twice for each model, one for function approximation and the other for testing and forecasting.

Model	Coefficient estimation method	Global error for function approximation (330 weeks)	Global error for testing (forecast-48weeks)
Multi Linear Regression	Least square	15.9	11.45
Vector autoregressive Analysis	Maximum likelihood	2.22	9.85
Neural Network$_{logsig}$	Gradient descent	7.01	30.69
Neural Network$_{radbas}$	Gradient descent	14.17	15.63
Neural Network$_{tansig}$	Gradient descent	6.71	25.62
Radial Basis$_{newrb}$	Least square optimization	5.33	29.25
Radial Basis$_{newrbe}$	Least square optimization	0.00	564.4
Radial Basis$_{newgrnn}$	Least square optimization	3.95	37.17

Table 29.Mean absolute percentage error (MAPE) of the different models

It is immature to set arbitrary forecasting targets; to say that a MAPE of <20% is accepted is not a common practice in forecasting, especially when forecasting volatile commodity prices that are governed by different supply and demand fundamentals. However, because the essence of this study is to evaluate and compare the forecasting power of different mathematical methods, the values of MAPE can be used.

As it can be seen in Table 29, the highest performance in function approximation is achieved in the maximum likelihood method as well as in the least squares optimization method. On the other side, most of the models exhibit a MAPE that is ≤30%, which, in statistical terms, is a good performance, especially that the analysis is conducted on time series that have more than 300 observations.

In order to illustrate the results, the residual error are computed and shown in Figure 18,

$$Residual\ error = \frac{Target\ price - Forecasted\ price}{Target\ price} \tag{26}$$

Figure 18 depicts the residual error for all models used in this study; i.e., the difference between the predicted values from the models constructed and the observed values are calculated for the period of January 6, 2014, to December 15, 2014.

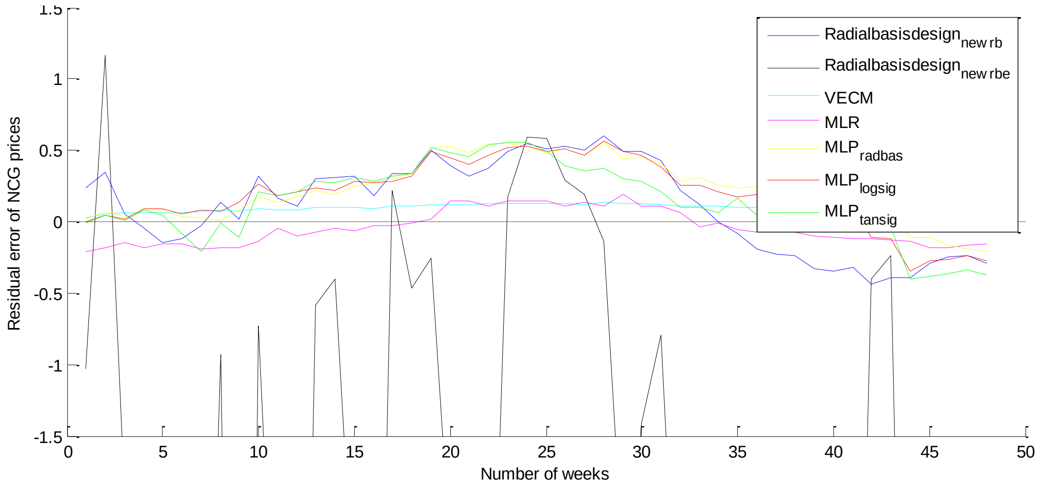

Figure 18. The variation of the values predicted by all models, from the observed values

The results are interpreted as follows:

i. The multi linear regression and the VAR imply good estimation with some relaxation of statistical assumption. This would force future modellers to rely on it with caution.

ii. Out of the remaining models, both the gradient descent and least squares optimization, which uses machine learning NN capabilities, succeed in having a low global error.

   It is also clear that, when it comes to function approximation, the gradient descent NN has a disadvantage over the models used in the least squares optimization. However, all three models of the former method ($\text{Neural Network}_{logsig}$, $\text{Neural Network}_{radbas}$ and the $\text{Neural Network}_{tansig}$) have consistent results when used on a new set of data.

iii. The least squares optimization is the technique that performs best in this study. The error results are the lowest for all three models in function approximation. However, one out of three models fails to give valid natural gas price forecasts.

   Consequently, the models that perform well in approximation use estimation techniques with strong capabilities of memorizing but poor performance when generalizing on a new set of data.

iv. Interestingly, the $newrb$ model has higher capabilities in terms of approximation and with good capabilities while fitting a new set of data, which is crucial in forecasting new short-term gas prices.

Chapter 7

7. Modeling the price of post liberalized gas markets – Information theory – Challenge 4

7.1 Nomenclature

X	…	Finite discrete random variable
n	…	Number of possible outcomes
x_i	…	A possible outcome
p_i	…	Probability of the possible outcome to appear
H	…	Shannon Entropy
r_t	…	Return of prices at time t
s_t	…	Discrete random variable – State
y_k	…	Discrete random variable – Signal
$p_{ik} = \mathbb{P}\left[{}^{S = s_i}/_{Y = y_k}\right]$	…	Conditional probability that the market is a certain state after receiving a certain signal
$\pi_i = \mathbb{P}[S = s_i]$	…	Probability of being in the state s_i
$H_k\left({}^{S}/_{y_k}\right)$	…	The conditional entropy of the random variable S relative to the signal y_k
q_k	…	Weight given for each signal
$H\left({}^{S}/_{Y}\right)$	…	"*posterior*" Entropy
$I(S, Y)$	…	Mutual information – Level and power of information

Table 30. Notation for the models applied in chapter 7

7.2 Introduction

Unlike the oil markets, the gas markets have witnessed regional divergence at several levels. However, the degree of competitiveness varies between the different gas markets.

Following extensive infrastructure development and regulation changes, transparent and competitive gas pricing hubs were developed in North America. Additional gas hubs have emerged afterward in Europe, providing physical and virtual locations for trading gas. The abundance of gas and the presence of competition between different stakeholders at different levels of the value chain have led to an increase in trade in the spot and the futures markets. However, the price of gas does not reflect market fundamentals and forces in all markets.

The liberalization of energy markets is an essential and fundamental policy tool used by authorities to regulate the natural gas sector. The role of the regulators is to promote competitive conduct, domestic gas production, third-party access, price trade reporting, and ensure the presence of futures trading. Once the measures are initiated, the status of the gas hub will be confirmed as liquid and stable, and the prices are considered as indicative of market fundamentals.

In this chapter, I focus on the North American and European markets, in specific the United Kingdom (UK), since both attempted to liberalize the gas markets, and underwent intense regulations and policy changes over the past years [125].

Wholesale buyers used to follow long-term contracts indexed to the price of oil derivatives in both of the aforementioned markets [50]. Also, the gas industry was mostly dominated by state-owned monopolies.

However, the Federal Energy Regulatory Commission (FERC), encouraged the establishment of gas markets which are driven by free competition in the United States [126]. As a result, the Henry Hub (HH), known as the most successful hub, was created [127]. The success of the HH is marked by a large liquid portfolio of spot and futures contracts, with hub indexed prices which serve as a reference for the value of the gas commodity all over the United States and North America.

Consequently, the UK and the European Union (EU) started their reforms. The Office of Gas and Electricity Markets (OFGEM) took lead and started the process of market liberalization since the 1990s. The reforms led to the establishment of the National Balancing Point (NBP), which serves as a physical platform for gas trading in the UK. Currently, the NBP is considered to be the most developed hub in Europe and has the longest standing European gas pricing point [128,129]. It is worth to mention that the UK gas market and the European gas market will be used interchangeably in the remainder of this study.

In 2016, the U.S. Natural gas consumption was roughly 750 Billion Cubic meters (BCM). The majority of the demand was satisfied through indigenous production, and the remaining was imported from Canada, via pipelines (see Figure 19).

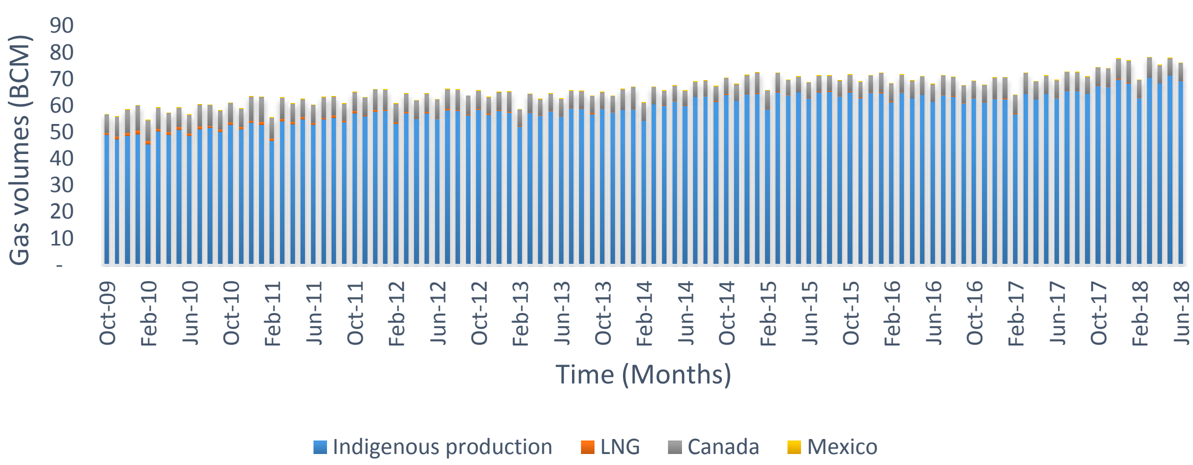

Figure 19. Indigenous production and monthly gas imports to U.S.[43]

[43] Source: EIA (Available on EIA website, https://www.eia.gov/naturalgas/monthly/).

The UK Natural gas consumption in 2016 is estimated at around 73 BCM, out of which 42 BCM are imported while the remaining volumes are produced locally (see Figure 20).

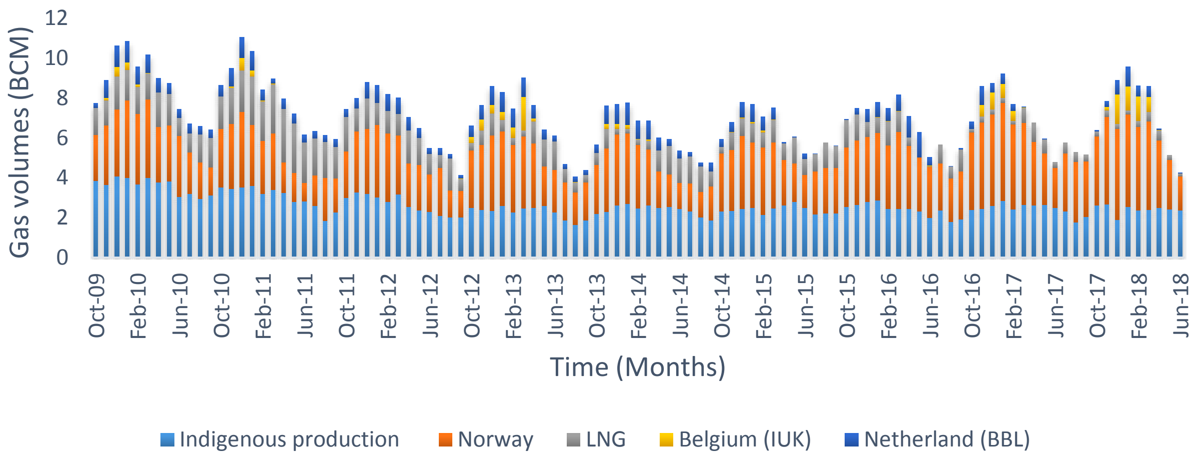

Figure 20. Indigenous production and monthly gas imports to UK[44]

To reflect the gradual advances in supply-side competition, the functioning of a wholesale gas market should be measured quantitatively, a subject that has attracted attention from the professional and scientific community, as in-depth analysis of gas markets have been conducted and published [20,29,30,50–52]. All studies confirm that parameters such as market participants, the monthly day ahead trades, and churn ratio give an indication and a feel of the market. The churn ratio, calculated as the ratio of traded gas volumes to the total gas demanded, is an indicative measure of the liquidity of a gas hub and market maturity, and it measures the confidence of traders and consumers in the market.

2015-2016				
Hub	**Market participants**	**Traded volumes, physical (BCM)**	**Traded volumes, Futures (BMC)**	**Churn ratio (Total)**
HH/U.S.	N/A	1,776	53,968	67
NBP/UK	40	901	925	20

Table 31. U.S. and UK traded volumes and churn ratio[45]

The numbers shown in Table 31, reveal a high churn ratio for both markets (above 15), and this indicates that the gas prices registered at both hubs are liquid and reflect market conditions [20]. Therefore, clearing prices

[44] Source: OFGEM –UK (Available on OFGEM website, https://www.ofgem.gov.uk/data-portal/gas-demand-and-supply-source-month-gb).
[45] Source: OFGEM [138], FERC [133] and Cornerstone Research, IEA.

are accepted as a reference and signal, which contain reliable information for all stakeholders involved in the gas value chain (traders, customers, regulators, etc.).

Analysis of recent trends in the European and North American gas markets shows that the prices of gas are fundamentally market-driven. However, rules and policies set by gas regulators are a must to guarantee that the market keeps on operating efficiently [130].

The first objective of this chapter is to study whether the wholesale gas prices of two of the most liberalized gas markets carry valuable information which can serve as signals for the relevant gas regulators. The value of these signals will be quantified by using several econometric methods and mathematical theories. This analysis will guide and assist the decision-making process of regulators regarding the need for an intervention to stabilize the gas markets and improve the functioning of their internal markets. The second objective is to quantify and measure the accuracy and efficiency of the hidden information structure generated by these signals.

All methods applied in this research are proven mathematical theories that have been used in previous studies. All four theories (Information, Records, Entropy, and Game theory) have applications in the field of mathematics and statistics, in other words, econometrics. However, the novelty of this method lies in adopting a two-step approach that was not applied before in the literature. This approach is useful to assess the performance of a gas market in terms of information generated by several signals and combined in one structure, called information structure. The signals give an idea of the level of competition, level of price volatility, and price stability on one hand, and the level of the information structure, which measures the performance of the market. Among all gas stakeholders, this information is important for gas regulators. These will be more confident and can trust the price signals if the performance of the market is powerful and efficient.

In the first step of the approach, I have identified three different signals describing market concentration, price stability, and price uncertainty of the considered market. Three appropriate mathematical and statistical theories are then applied to extract, from the gas prices time-series metric values that are most suitable to measure the relevant signals. The formulation of each model is explained and justified in the next section.

The level of competition changes from one market to another, and if measured correctly defines the concentration of competing firms in the market. Few and a limited number of firms imply a highly concentrated market, and that based on their strategies can dictate prices, otherwise known as price-setters. Besides the fewer the number of firms the easier it is to abuse conduct and act collusively. Such firms adjust their strategies in conjunction with an agreed-upon understanding with the competing firms at the expense of the welfare of gas consumers and possibly smaller firms. A typical example of such a market and behavior is the presence of cartels in commodity markets.

The level of volatility indicates how fast and sudden gas prices change in the short term. The higher the volatility the harder it is to predict the future behavior of the changes, thus making the market uncertain.

Price stability hint at the behavior of gas prices in the medium and long term on the other hand. Commodity prices tend to have abrupt and rapid price shocks, and this is witnessed when prices suddenly increase or decrease due to sudden changes in supply and demand characteristics. The longer it takes for a commodity price to witness a shock the more stable the market is.

In the second step, the three signals are combined to create an information structure that will help the authors evaluate the performance of the gas market in question. The performance of the market is the measure of the power and efficiency of the information contained in the gas prices and the signals. The more efficient, the more reliable, and reflective the prices are in such a market. The signals are assessed against the actions/ states that could be executed by the regulator of such markets. Two actions are identified; either to intervene in the market by taking legal actions, such as issuing new directives in order to ensure a stable supply and demand equilibrium and making sure that there are no abusive conduct by gas suppliers; or not to intervene. Furthermore, the approach deals with a case where the information is neither completely absent nor perfectly known, which has rarely been dealt with in literature.

7.3 Material and methods

To avoid price abuse and manipulation, a gas regulator is expected to regulate firms' behavior by ensuring that customer welfare is maximized while maintaining the attractiveness and profitability for the producers and traders.

As stated in the introduction, price dynamics of a commodity in a liberalized market are indicative of the market structure and contain consistent information that should, if adequately analyzed, help the regulators in assessing the performance of the market, namely the wholesale gas market in this case.

The authors have identified three main metrics that can signal information in the hidden structure of the price values of both hubs. These metrics are based on econometric and mathematical methods, and are used to inform the regulator in each market about the following:

- **Signal 1: Level of competition**
- **Signal 2: Market stability**
- **Signal 3: Volatility and uncertainty of prices**

The first signal studies the degree of concentration in the two different gas markets by using Game Theory, specifically the non-parametric Nash-Cournot equilibrium test. In other words, if the test shows that traders are participating in the market by trying to maximize their profit as "the only pure" strategy, then the market is considered efficient and the likelihood of anti-competitive behavior is negligible.

The second signal employs the Records Theory, which relies on the analysis of the peak observations reached in a certain period, and that exceeds the previous observations. This signal measures the degree of market stability, by calculating the probability of witnessing future peak prices. Therefore, the measure of probability is a measure of market stability and predictability. If the results point towards a tendency to score high probabilities of extreme gas prices, then the market can be characterized as unstable.

The third and final signal studies the price predictability of both markets, by the use of Shannon Entropy and the measure of volatility. This is done by analyzing the variation of prices and returns and assessing the degree of uncertainty and volatility which are present in gas prices. Simply, the higher the uncertainty in prices, the higher the volatility.

These signals combined will inform the regulator about the functioning of the market. If the market shows signs of concentration, the likelihood of extreme prices, volatility, and uncertainty, then the regulator should intervene and use its policy enforcement power.

Since the signals are based on econometric theory and models, it is important to assess the performance of such models. Therefore, a quantitative analysis that relies on Information Theory is used to compare the power and efficiency of the information generated by all signals in the two different selected markets. The market with the highest information power will give additional credibility to the signals so that the price signal speaks for itself. Regulators in such a market have higher confidence and can trust the signals, which will guide their decision of whether to intervene in the market or not.

The following part of section 7.3 will list and define the three signals metrics and will explain the econometric and mathematical methods that will be used in this chapter. Sub-section 7.3.5 will lay out an outline of the power of information. Results will be presented in section 7.4 with an analysis of their significance and impact, with an overview of how they can be used by gas regulators in their assessment of wholesale gas markets.

7.3.1 Data

A description of the data used in the study is presented in Table 32. The variables set consist of monthly wholesale gas prices that were registered between October 2009 and June 2018.

Variable	Frequency	Number of observations	Source
The National Balancing point, NBP gas prices (2009-2018)	Monthly	105	National Regulatory Authority – OFGEM Office of Gas and Electricity Markets – UK Government[46]
The Henry Hub, HH gas prices (2009-2018)	Monthly	105	U.S. Energy Information Administration (EIA)[47]

Table 32.Data collection

Figure 21 illustrates, in a time series plot, the evolution of the natural gas prices for the two different markets. There is a clear divergence that happened in the year 2009 and continues to date. This is mainly due to two main factors that took place in the U.S. The first is the abundance and oversupply of new unconventional shale

[46] Available on OFGEM website, https://www.ofgem.gov.uk/data-portal/gas-prices-day-ahead-contracts-monthly-average-gb.

[47] Available on the U.S. Energy Information Administration (EIA) website, https://www.eia.gov/dnav/ng/hist_xls/RNGWHHDd.xls

gas production in the local market. The second is that U.S. natural gas contracts in that period started to be decoupled from the crude oil prices.

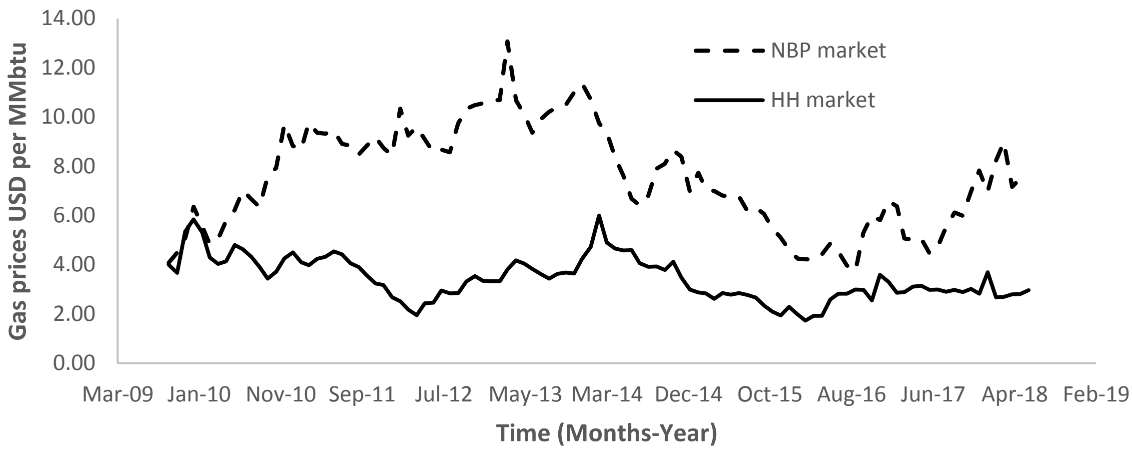

Figure 21. Monthly Gas Prices of both Gas markets, USD/MMBtu

The line plots of the two markets presented in Figure 21 show that there is no clear indication of a linear relationship between both variables.

7.3.2 Signal 1 – Level of competition

The level of competition and market concentration method involves the classical Nash-Cournot equilibrium test. A Cournot equilibrium is reached when a given firm maximizes its profit by changing its output taking into account the other firm's output. One important feature of a Cournot model is that firms are not allowed to cooperate. Therefore, as long as the players are playing the latter strategy, the companies would be abiding by pure market profit-oriented strategies, trying to maximize their 'utility function', and have no agreed-upon behavior (i.e. collusion). The market, where producers follow this trend is considered more liberalized.

The aim is to test whether the behavior of the gas producers in the respective markets follows a Cournot model. If a set of gas producers are not following the assumptions of a Cournot game, the test will identify them. To conduct the problem of detection of Cournot's equilibrium, a numerical algorithm was developed. The non-parametric algorithm is explained in details in chapter 5. The result of which indicates if the data being tested respects or not the Cournot equilibrium. I invite the reader to review chapter 5, more specifically both conditions (1 and 2) of the algorithm, as well as the Cournot acceptance rate.

In simple words, the algorithm will compute the Cournot acceptance rate for each of the markets that will be tested. There are three expected outcomes:

1- Companies are competing based on a Cournot model, trying to maximize their profit by acting strategically. In this case, Cournot acceptance is high.

2- Companies are cooperating and not acting by the rules of Nash Cournot equilibrium. In this case, Cournot acceptance is low.

3- Other strategies and objectives can be set by competing firms, however, the Cournot acceptance rate is only used to differentiate if the latter companies are respecting conditions 1 and 2, which are linked to the Nash-Cournot model.

7.3.3 Signal 2 – Market stability

Records theory studies observations that are concentrated in the tail of a given distribution [15] and will be used in this context to test the stability of two different gas markets. The Theory of Records is explained in details in chapter 4.

The aim of this analysis is to compute the probability of witnessing a price record (peak observation) in the long term for each of the gas markets in question. The closer the probability is to zero, the better (in terms of low risk of sudden changes in gas prices) it is for all participants in the gas value chain in such a market.

7.3.4 Signal 3 – Volatility and uncertainty in prices

The third quantitative method used in this study is represented by Shannon's Probabilistic Entropy and is used on a time series analysis, to test the predictability power hidden in the underlying probabilistic distribution of the considered time series. A time series with a high predictability power is considered to have a high level of stability with an anticipated pattern.

The classical definition of entropy is as follows: for a given source of information represented by a finite discrete random variable X with *n* possible outcomes, each possible outcome x_i having a probability p_i to appear, the Shannon entropy H of the random variable X is defined by:

$$H(X) = -\sum_{i=1}^{n} p_i \, log_2 \, p_i, \quad (1)$$

In general, a logarithm of base 2 ($\log_2$) is used because the entropy is generally expressed in bits [93]. Several researchers have previously attempted to predict the entropy of the commodity markets (oil, more specifically Brent and West Texas Intermediate, WTI, and other commodities) and tried to measure the information from statistical observations [89–91]. Brent and WTI are two different crude oil grades (quality) and are known to be the most important oil pricing benchmark around the globe. As previously explained, the gas markets in question have been liberalized, and the influence of oil prices on gas prices is shrinking. Gas prices are becoming more influenced by gas to gas competition. To the knowledge of the authors, no previous researchers have worked on predicting the entropy of the gas markets.

To compute the Shannon entropy of a time series, which is a continuous random variable, a particular discretization method will be introduced:

First, the returns of the prices are computed, this is a requirement for normalizing the data set

$$r_t = 100 * \frac{P_t - P_{t-1}}{P_{t-1}}, \tag{2}$$

where P_t and P_{t-1} are the prices at times t and $t-1$ respectively.

It's trivial that the series of observations of returns r_t has an underlying continuous distribution. Therefore, the second step is to introduce the discrete random variable s_t defined by:

$$s_t = \begin{cases} 1 & \text{if } \ r_t \geq 0 \\ 0 & \text{if } \ r_t < 0 \end{cases} \tag{3}$$

The random variable has a binary output. Unity is when the returns are positive, which means that the prices are increasing, and zero signals that the prices tend to diminish. Based on the observed values of s_t, and by denoting the total number of observations by n, the corresponding probability distribution is computed:

$$p_1 = \mathbb{P}[s_t = 1] = \frac{\sum_{t=1}^{n} s_t}{n}, \tag{4}$$

and,

$$p_0 = \mathbb{P}[s_t = 0] = 1 - p_1 \tag{5}$$

Hence, the entropy related to the random variable s_t is given by:

$$H\ (s_t) = \ -p_0 * log_2(p_0) - p_1 * log_2(p_1) \tag{6}$$

To compute the underlying entropy of each gas market by the above-explained method, I rely on daily values instead of monthly values. The price variable is divided into year windows, with 252 observations for each. The passage from one window to the other is done by removing the first observation while adding another from the remaining ones, and so on. By applying the latter procedure, I obtain a series of entropies that should be represented by the mean parameter, as a representative of a series of entropy observations.

However, a major disadvantage of the mean is that it's sensitive to distribution with a thick queue which can be caused by the presence of outliers and extreme observations. Besides, the mean may have a false interpretation in case of a highly skewed distribution for the considered data. To overcome these weaknesses, a second statistical parameter will be adopted: the median value. Moreover, it has been shown that the median is useful when comparing sets of data. Once the Median entropy of each market is computed, the values for each market should be compared. In addition, a non-parametric mathematical test will be conducted to check the statistical significance of the difference between both markets [131].

The Kruskal test is used to compare two independent samples and checks if the observations originate from two different distributions or not.

Finally, the volatility for each market is computed and can serve to validate the results of the price unpredictability. It is the degree of variation in the price series of each market and is measured by the classical standard deviation of the returns.

Note that one can find in literature a version of the entropy adapted to the continuous random variable cases, called differential entropy [132], given by:

$$H(X) = -\int_{x} f(x) * \log_2 f(x)\ \mathrm{d}x \tag{7}$$

Where $f(\cdot)$ is the probability density function of the underlying distribution of the continuous random variable X. However, this method has many flaws:

- The density function $f(\cdot)$ is in general unknown. This is a weakness because the users will make assumptions about the distribution type. Nonetheless, users can utilize numerical methods to estimate the density function empirically, however, one could face several challenges concerning errors of estimation.

- The properties and the interpretation of discrete random variables entropies are not known to be conserved when passing to the continuous case. In other words, the differential entropy does not share all properties of discrete entropy.

7.3.5 Information theory

After defining the signals, which can be extracted from the gas prices of each market, the regulator has to make important decisions. If the market signals indicate market concentration and price instability, then certain measures should be taken to bring back stability to the gas prices. Therefore, the two defined states in this study are either for the regulators to take action or keep the business as usual (BAU). This is defined by $s_i, i = 1{,}2$ respectively. The signals previously defined will be denoted by $y_k, k = 1{,}2{,}3$ respectively. Also, I denote by p_{ij} the conditional probability that the market is in state s_i after receiving the signal y_k i.e.

$$p_{ik} = \mathbb{P}\left[{}^{S = s_i}/_{Y = y_k}\right] \tag{8}$$

The probabilities p_{ik} are categorized into three classes: Low, Medium, and High. Each class has the following respective probabilities: $\frac{1}{10}, \frac{1}{2}\ and\ \frac{9}{10}$. The information structure is illustrated in Table 33.

The methods used in this paper reduce the subjectivity of the probability distribution. The signals are complemented by econometric models founded by economic parameters of the relevant gas markets, and by data analysis on its gas prices, from which the information will be extracted.

The probability of being in a certain set (either s_1 or s_2) after receiving the signal (either y_1, y_2, or y_3) can take either a value of 0.1, 0.5, or 0.9, which constitutes the possible events on the probability set. The sum of the probability of being in s_1 or s_2 after receiving the same signal y_1, however, should be equal to 1. This is normal as there are only two sets considered in this study.

$$\sum_{i=1}^{2} p_{i1} = p_{11} = \mathbb{P}\left[{}^{S = s_1}/_{Y = y_1}\right] + p_{21} = \mathbb{P}\left[{}^{S = s_2}/_{Y = y_1}\right] = 1 \tag{9}$$

States	y_1- Level of competition	y_2- Market stability	y_3- Volatility and uncertainty of prices
s_1- Action needed	p_{11}	p_{12}	p_{13}
s_2- No action (BAU)	p_{21}	p_{22}	p_{23}

Table 33. Information structure conditional probabilities.

Note that the first step is to compute entropy, called "apriory" entropy, based on the distribution of the states S before the reception of any additional information. Then, I start by:

$$H(S) = -\sum_{i=1}^{2} \pi_i \log_2 \pi_i \tag{10}$$

Where $\pi_i = \mathbb{P}[S = s_i]$ is the probability of being in the state s_i before receiving additional information called "*apriory probability*". As π_i is defined based on no previous information, it is reasonable to consider a distribution, which has the highest level of uncertainty. In other words, the regulator has no information that can lead him to make an action, and the probability of either of the two states is equally likely. This is a uniform distribution with the following "*apriory probabilities*" $\left(\pi_1 = \pi_2 = \frac{1}{2}\right)$.

Now, after receiving a specific signal of information y_k, the conditional entropy of the random variable S relative to the signal y_k is defined as:

$$H_k\left({}^{S}/_{y_k}\right) = -\sum_{i=1}^{2} p_{ik} \log_2 p_{ik} \tag{11}$$

Hence, to assess the power of information generated by the whole information structure (composed by y_1, y_2, and y_3), the "*posterior entropy*" is defined, and is compared to the "*apriory probability*" :

$$H\left({}^{S}/_{Y}\right) = \sum_{k=1}^{3} q_k * H_k\left({}^{S}/_{y_k}\right) \tag{12}$$

Where q_k is the weight given for each signal. This number is equally distributed for the three signals, as all three signals are equally important and each contributes to the understanding of the gas market in different ways.

Finally, the amount of reduced uncertainty, due to the additional received information, is measured by a quantity called mutual information:

$$I(S,Y) = H(S) - H\left({}^{S}/_{Y}\right) \tag{13}$$

Thus, to compare the power of information generated by two information structure related to two different gas market, one should consider the one with the highest $I(S,Y)$. This is equivalent to say that the structure of information that reduces most of the uncertainty of the random variable S will be most efficient and powerful.

7.4 Results and discussion

7.4.1 Results of the non - parametric Cournot test

Considering Table 31, it is evident that both gas markets are competitive. Nonetheless, this is considerable and significant in the case of HH. Table 31 draws attention to two main numbers: the first being the big difference between the volumes of gas traded in the future and the volume traded on the physical, which indicates the excessive participation for traders and financial players in the virtual market. The second being the large numbers of churn ratios, which indicates high liquidity and healthy trading platform, an attractive characteristic for all stakeholders. Unlike the U.S. gas market, the Herfindahl index for the European gas market is relatively high [20]. This is a sign of healthy competition, and this simply means that out of the many gas suppliers in the U.S. market, none has market power on its own. However, this violates one of the main assumptions of a Cournot competition model, where firms have market power, and each firm's output decision affect the gas prices. In a nutshell, there is no risk of market manipulation in such a market, therefore the market concentration is minimal and close to zero. All U.S. gas suppliers should be price takers in this case, and the Cournot acceptance rate is no longer valid.

The data that is used for the Cournot test consist of the gas prices and the gas supplies to the relevant market. Gas supplies are shown in Figure 19 and Figure 20. The suppliers are represented by countries of origin. Results can be more indicative if the data related to gas supplies is composed of volumes of the suppliers (shippers, trader, and companies) directly rather than the country (market) where the gas was purchased. The authors acknowledge the need for the traders' suppliers' data and the need to perform the Cournot test on the American market, however with no publically available information on the supply market shares of companies, this is not possible. Therefore, I encourage the publishing agencies to list such data on their website (or upon request). The FERC Form 552 provides a database of trading activity and lists the data related to the largest companies (Top 20) with the largest total transaction volume from year to year. The list found in [133] is incomplete and contains yearly data only. Thus additional data related to suppliers' portfolios is needed to have valid test results. The suppliers in North Western Europe are oligopolistic [53]; therefore, the usage of the data will lead to conclusive and significant results, when using the algorithm.

The Non- parametric test results for the European market gives a Cournot acceptance rate of 51%. The results can be analyzed as follows: the behavior of the large gas suppliers in the European can be explained by a Cournot model, where suppliers are trying to maximize their payoffs by competing over quantities. However, the other half of the acceptance rate means that there are companies that are not behaving as such. This could implicate that some of the suppliers have other strategies such as collusion, or strategies that are not "pure" profit maximizers.

An example of a possible collusive behavior has been witnessed in the oil markets under the Organization of the Petroleum Exporting Countries, OPEC back in the 1970s. These countries use to control a major share of the world oil supplies, and together they form a cartel that cooperates, with the aim to increase prices and limit external competition.

Other examples that can be used to illustrate possible reasons why these suppliers are not seeking a "profit only" strategy under the Nash-Cournot umbrella are listed below:

Authors such as [37], suggest that Gazprom, a major gas suppliers, is maximizing its 'utility function' not only by limiting itself on one strategy that is focused on making a profit, but also contemplating other strategies such as seeking to eliminate competition, even if this leads to some losses in profits initially.

Other authors, such as [36], enumerate other reasons that are preventing some of the European gas suppliers from exerting their oligopolistic power, and these are due to old legacy contracts that are still effective, and perhaps new regulations. In short gas prices mechanism in old legacy contracts are mainly indexed to oil prices, and this type of contract does not offer the needed flexibility to gas suppliers. These valid assumptions are among many, possible reasons why the Cournot acceptance rate is not that elevated in Europe.

The first signal is informative and the analysis of the prices of the NBP wholesale gas prices is indicative for the regulator in this market.

7.4.2 Results of the Records theory

The second signal is assessed using the Records Theory. To anticipate if the data belongs to an $i.i.d$ sequence of variables, the goodness of fit test is used.

Markets	HH	NBP
p-value	0.89	0.05
Result	Accept H_0	Reject H_0

Table 34. Goodness of fit test results

The results shown in Table 34 indicate that the European market rejects the null hypothesis. The test results were computed at a confidence level of 5%. Accordingly, and from an empirical perspective, the gas prices recorded in this market are characterized by price variations and sudden price shifts.

Based on the analysis of Table 35, the result is not surprising, as it confirms that the European market has a high number of records relative to the small number of observations. This indicates that the European gas price records are not grouped in one section of the time series, and are instead more spread, while the U.S. market is rather more stable and that price shifts are rarely observed all along the time series.

Markets	HH	NBP
Number of Records	_2_	_8_
Record index	Nov-09 Dec-09	Nov-09 Dec-09 Jan-10 Oct-10 Dec-10 Feb-12 Mar-13 Apr-13

Table 35. Number of records and record index

Looking further, in an attempt to measure the probability of witnessing a record in each of the gas markets, the Yang model will be used for the European market and the classical model for the U.S. market. The computed probabilities were computed for each market, and Table 36 shows the result of the probability that matches the date of June 2018.

Markets	**Probability of Records**	**Probability of having a record on $t = 105$ (June 2018)**
HH	$\mathbb{P}[\delta_t = 1] = \mathrm{P_t} = \frac{1}{t}$	$\mathbb{P}[\delta_t = 1] = 0.01$
NBP, where $\boldsymbol{\gamma}$ is equal to 1.016	$\mathbb{P}[\delta_t = 1] = \mathrm{P_t}(\gamma) = (\gamma^t(\gamma - 1))/(\gamma(\gamma^t - 1))$	$\mathbb{P}[\delta_t = 1] = 0.02$

Table 36. Probability of Records results

The probability of witnessing a new record is higher in the European market. The results from the above analysis can be summarized as follows:

1) The record rate in the European model converges to a constant value in the medium and long term (Typical for Yang model, see section 4.3.4). This is an indicator that the market could prove to be unstable over time.

2) In contrast, the record rate and the time index in the U.S. model tend to have a negative correlation. This means that the probability of record diminishes over time, i.e. when t increases, the probability of record diminish. This is an indicator that the market is rather more stable in the medium and long term.

7.4.3 Results of the Shanon entropy

By applying the procedure described in Section 7.3.4 dealing with Shannon Entropy, the representative median entropy of each considered market in addition to the p-value of the Kruskal non-parametric test is calculated. The results of the entropy approach are presented in Table 37:

Markets	HH	NBP
Median Entropy	0.9939	0.9991
Kruskal p-value	2.2e-16	

Table 37. Entropy results

If a random variable X follows a discrete uniform distribution with n possible outcomes, the corresponding entropy is $H(X) = log_2 n$ [134]. Hence, in our context, the values of the entropies are both close to the case of uniform distribution $log_2 2$, which is equal to unity, and this means that both markets are far from being predictable.

As the values of the median entropies of the considered two markets are close to each other, it is substantial to test if the two considered median entropies are issued from two different distributions. If it is the case, then this indicates a significant difference between the two medians. The non-parametric Kruskal test is applied to verify the latter hypothesis.

Based on Table 37, the p-value of the Kruskal test is close to zero, and therefore less than 5%. Accordingly, the difference between the markets in terms of entropy is significant, i.e. the market with the higher median value, European market in our case, has an entropy significantly higher than the U.S. market.

Also, the volatility in the U.S. market is very low (0.7), whereas it is significantly high in the European market (2.5). Thus, for signal number three, the U.S. market is significantly more predictable and has lower uncertainty than the European market.

7.4.4 Synthesis of signal results

In an attempt to better illustrate the results of the three mathematical models used in the previous sections to measure the market signals, Table 38 and Table 39 summarize the results and list the main findings for each market.

y_1- Level of competition	y_2- Market stability	y_3- Volatility and uncertainty of prices
The Cournot acceptance rate is calculated: • $\delta = 51\%$ This number gives a good indication that the oligopolistic gas suppliers in Europe are playing a Cournot game, where the supplier/ company is trying to maximize its payoff by competing over quantities. The Cournot acceptance rate also indicates not all the companies are acting as profit maximizers and playing a Cournot game. In fact, less than half of the observations do not respect conditions 1 & 2, which means, companies are not respecting the rules of a Nash-Cournot model. This indicates that some companies are either adopting strategies that are not "pure" profit maximizers on one hand or that they are colluding with the aim to increase profit on the other hand. This is unacceptable and thus will reduce consumer welfare.	As per the results of the Goodness of fit test (Table 34), the EU gas prices behave in a non-$i.i.d$ case. The record rate calculated for the Yang model converges to a constant value in the long term. The results of the Yang model are as follows: • $\mathbb{P}[\delta_t = 1] = 0.02$, for t=105, equivalent to the month of June 2018 • $\mathbb{P}[\delta_t = 1] = 0.02$, for t=1100, equivalent to the month of May 2100 The probability of records converges to constant in the long term, which means that there is always a possibility for sudden and drastic change (spike or drop) in gas prices.	The entropy for the EU market is close to a uniform distribution. • The median entropy for the NBP market is 0.9991 • The volatility of the NBP market is relatively higher and is at 2.5 As for the entropy of the U.S. market, its entropy is also high. Thus, both markets are close to the uniform distribution, which means that both are far from being predictable. The Kruskal non-parametric test result is: • p-value is 2.2e-16 This implies that, relative to each other, the entropy of the NBP gas market is significantly higher than the HH gas market. Thus high volatility and higher median entropy imply that the European gas prices have bigger uncertainty in the medium term and thus not easily predictable.

Table 38. Synthesis of signal results for the EU market

y_1- Level of competition	y_2- Market stability	y_3- Volatility and uncertainty of prices
As mentioned in section 7.4.1, the U.S. market is characterized by: • The number of traders participating in the market is high. • A large number of churn ratio • Herfindahl index compared to the European gas hubs is low This simply means, that unlike the European market, no single supplier trading in the U.S. has sufficient market power. In this case, the Cournot acceptance rate is not valid, since This violates one of the main assumptions of a Cournot competition model, where firms have market power, and each firm's output decision affect the gas prices. • $\delta = \mathrm{N/A}$	As per the results of the Goodness of fit test (Table 34), the U.S. market is modeled as the case of a classical record. The results of the classical model are as follows: • $\mathbb{P}[\delta_t = 1] = 0.01$, for t=105, equivalent to the month of June 2018 • $\mathbb{P}[\delta_t = 1] = 0.00$, for t=1100, equivalent to the month of May 2100 The Probability of records converges to zero in the long term. This simply means that the market is stable, and there is no risk of witnessing sudden price changes.	The entropy for the U.S. market is also close to a uniform distribution. • The median entropy for the HH market is 0.9939. • The volatility of the HH market is very low and is at 0.7. Thus low volatility and lower median entropy imply that the U.S. gas prices have smaller uncertainty in the medium term and easier to predict.

Table 39. Synthesis of signal results for the U.S. market

7.4.5 Results of the Information theory

Concentrated markets raise regulatory and antitrust concerns, as this is a clear sign of market power in the hands of suppliers. Appropriate actions need to be taken by the regulator to make sure that neither collusion, nor cooperation between companies, nor any kind of strategic decisions that do not end up in favor of consumer welfare, are permitted.

The regulator in such a case should ensure that under no circumstances, the companies communicate and have the agreed-upon understanding to raise prices and profit margins at the expense of consumer welfare. Barrier to entry for new companies should also be considered and reduced by regulators in such markets, to increase competition and diversify supplies. These are some examples of actions that the regulator can impose on the suppliers.

Markets that witness price volatility and uncertainly in the medium term, as well as price instability in the long term, are also raising concerns for regulators. In such a case the key to determining the movement of gas prices are the supply and demand fundamentals. A slowdown in global demand is a key downside risk for suppliers, as they will eventually earn less while trying to sell their gas. On another hand, a sudden slowdown in supply is a key downside risk for another player in the gas value chain, which is the consumer. The latter will have to pay more to purchase the commodity.

In both cases, regulators should anticipate such results by acting in favor of a continuous supply and demand equilibrium, by trying to diversify supply (indigenous production, imports, and storage), while also ensuring that the consumers have the appropriate infrastructure and financial means to buy such a commodity. However in a market characterized by gas prices that are predictable in the medium term and prices that are stable over the long term, then there is no need for further actions by the regulator.

Moving forward, I start by assigning the relevant conditional probabilities p_{ik} which indicate to the regulator the state of nature of the gas market. As previously mentioned, the probabilities are categorized into three classes: Low, Medium, and High. Also important to remember that the sum of $\sum_{i=1}^{2} p_{i1}$is equal to 1, as there are only two possible sets. The same thing applies to signals 2 and 3.

After presenting the results of the three signals in sections 7.4.1 to 7.4.4, Table 40 and Table 41 explain the process of probability category selection and list the results for both the U.S. and European markets.

The results listed in Table 40, give an indication that the regulator in the European market is more inclined to intervene. UK's regulator OFGEM has to intervene and investigate the reason behind some instability and signs of non-competitive behavior, where some firms are not only focused on profit maximization.

EU market – NBP			
State	y_1- Level of competition	y_2- Market stability	y_3- Volatility and uncertainty of prices
s_1- Action needed	High Cournot acceptance rate, implies that the large gas suppliers are trying to maximize their payoffs by competing over quantities. However, the other half of the acceptance rate means that there are companies that are not behaving as such. A possible indication of market abuse, therefore the regulator in invited to intervene. **MEDIUM**, $p_{11,\mathrm{EU}} = \mathbf{0.5}$	High record probability that does not converge to zero on a long term, implies that the market is not stable and there is a need for additional market oversight measures. **MEDIUM**, $p_{12,\mathrm{EU}} = \mathbf{0.5}$	High volatility and higher entropy values (relative to the U.S. market) means that the gas prices are not predictable. **HIGH**, $p_{13,\mathrm{EU}} = \mathbf{0.9}$
s_2- No action (BAU)	The regulator needs to intervene, to adjust the legal framework along with the supply and demand fundamentals. This is essential as it ensures a smooth functioning gas market. **MEDIUM**, $p_{21,\mathrm{U.S.}} = 1 - p_{11,\mathrm{EU}} = \mathbf{0.5}$	The regulator has to take action. **MEDIUM**, $p_{22,\mathrm{U.S.}} = 1 - p_{12,\mathrm{EU}} = \mathbf{0.5}$	The regulator must take action. **LOW**, $p_{23,\mathrm{U.S.}} = 1 - p_{13,\mathrm{EU}} = \mathbf{0.1}$

Table 40. Information structure conditional probabilities for the EU market

The results listed in Table 41, give a clear indication that the market in the U.S. is functioning smoothly and that the regulator does not need to add other measures. In other words, the BAU case is favored.

U.S. market - HH			
State	y_1- Level of competition	y_2- Market stability	y_3- Volatility and uncertainty of prices
s_1- Action needed	A liberalized and non-concentrated market, therefore the probability assigned for any action by the relevant regulator is minimal. **LOW**, $p_{11,\text{U.S.}} = \mathbf{0.1}$	Low record probability that converges to zero on a long term, implies that the market is stable and there is no need for additional market oversight measures. **LOW**, $p_{12,\text{U.S.}} = \mathbf{0.1}$	Low volatility and lower entropy values (relative to the European market) means that the gas prices are far from being unpredictable, although the entropy values are not that low when analyzed with no benchmark. **MEDIUM**, $p_{13,\text{U.S.}} = \mathbf{0.5}$
s_2- No action (BAU)	The regulator does not need to intervene, and the legal framework along with the supply and demand fundamentals are ensuring a smooth functioning market. **HIGH**, $p_{21,\text{U.S.}} = 1 - p_{11,\text{U.S.}} = \mathbf{0.9}$	No intervention needed. **HIGH**, $p_{22,\text{U.S.}} = 1 - p_{12,\text{U.S.}} = \mathbf{0.9}$	No intervention needed. **MEDIUM**, $p_{23,\text{U.S.}} = 1 - p_{13,\text{U.S.}} = \mathbf{0.5}$

Table 41. Information structure conditional probabilities for the U.S. market

In order to compute the global power of information generated by the considered information structure, I start by assessing the level of uncertainty of each receiving signal by computing the conditional entropy of the latter $H_k\left(S/y_k\right)$; then I get:

Market	y_1- Level of competition	y_2- Market stability	y_3- Volatility and uncertainty of prices
U.S. – HH	0.47	0.47	1
Europe – NBP	1	1	0.47

Table 42. Conditional entropy of each signal

The "*posterior entropy*", previously defined by $H\left({}^{S}/_{Y}\right)$ is then computed, and compared with the "*apriori entropy"*, which is defined in section 7.3.5 as the entropy of a uniform distribution (one that has the highest level of uncertainty), and given a value of 1.

Market	$Posterior$	$Apriori$
U.S. – HH	0.64	1
Europe – NBP	0.82	1

Table 43. "Posterior" and "apriori" entropies

Table 42 and Table 43, illustrate the results of the entropies, conditional to the relevant signals, which is then used to compute the outcome of these signals in aggregated dimension and for each market.

The difference between the "*posterior entropy*" and the "*apriori entropy"* will help assess the level and amount of information, previously defined as $I(S,Y)$ gained by analyzing the gas prices data in each market. In other words, the signal analyses that is measured by the various econometric methods used in this study, constitutes an additional information that the regulators can use to assess the status of the market. The more the additional information increases (i.e. the difference between the "*posterior entropy*" and the "*apriori entropy")*, the more confident the regulator is about the power of information generated.

The amount of reduced uncertainty, due to the additional information received from the signals, is estimated at 0.38 for the U.S. market and 0.18 for the European market, which means that the level of uncertainty has been reduced in the European and U.S. market respectively by 18% and 38%. The value of the information contained in both markets, although in asymmetric terms, is significant, powerful and can serve as a reliable and efficient source of information.

Chapter 8

8. Synthesis of results and conclusions

8.1 Findings, discussions and conclusions referring to the research questions

In an attempt to answer the four set of challenges and research questions related to the gas market liberalization trends in several regions listed in Chapter 1 (Section 1.2), various econometric models have been developed. Chapters 4, 5, 6 and 7 have touched on the methods developed, the data used and the results.

This section will restate the four main challenges and summarize the main highlights of the work, in an attempt to answer the research questions.

Challenge 1 – The nominal "one price" for Gas markets

To date the literature studying the tails of the distribution, have used the extreme value theory to study the price behavior of commodity prices. The research presented in chapter 4; however offers an alternative and concentrates on the records theory.

The findings suggest that out of the three main regional gas markets, the Asian market seems to be less stable than the others, and that the probability of having a record in the coming years is the highest. In addition, there is always a probability of having a record in the long run for the Asian market, which proves that the market is not stable, unlike the American and European markets which seems to be more stable in both the short and long run.

While this study offers binding mathematical models, it does also reveal concrete results that are distribution free. Most importantly the validity of the models/ results is validated in the two step analysis I have conducted, as both the empirical and theoretical findings seems to be close for all three models, especially the Asian market.

The analytical framework that is based on distinctive mathematical models described in this thesis, and the ability to forecast future spike / drops should be used as an incentive, by Asian gas regulators, where there are signs of price instability.

While the private sector is mainly responsible for commercial deals, the public sector is encouraged to support and improve business environment for developing well-working rules of gas markets conductive to a better functioning LNG market, especially in terms of flexibility, price formation, gas supply security, and securing necessary investment.

Answer to Challenge 1*:* In the short term, the global gas prices will not head to a nominal “one price”, and this is mainly due to the instability of the gas market in Asia and partly in Europe. Reforms in terms of legal framework, technology advance in terms of LNG transport and more investment are three main key areas that the public and private sector should be focusing on in such markets.

Challenge 2 – European Gas price mechanism

This challenge presents a novel approach to testing if the liberalization of the gas industry has led to less concentrated European gas markets and studies the behavior of gas suppliers. The study of gas markets in itself is not new; previous analyses follow classical parametric optimization methods, which focus on the use of empirical data and market assumptions about cost and demand functions, for this purpose. By contrast, the method used in this study requires no parametric assumptions about demand and cost functions.

Using the non-parametric method, I assess the degree of concentration in two different European gas markets, Austria and the Netherlands, where each represents a different evolutionary stage in the process of wholesale gas market liberalization.

The results show that the gas suppliers’ behavior in both the Dutch and Austrian markets follows the Nash-Cournot equilibrium and that they are rationally acting to maximize their payoffs. More importantly, it validates the fact that suppliers in both markets engage in the same conduct even though one market is more liquid than the other. The results also show that not all the observations $\mathcal{C}$ respect the Cournot equilibrium, which means that the suppliers in that market are maximizing their ‘utility function’ not only by seeking profit but also by pursuing non-profit objectives, such as cooperative collusive behavior. This means that some suppliers are cooperating and thus exhibiting non-competitive behavior. Such suppliers adjust their strategies in conjunction with an agreed-upon understanding with the competing suppliers at the expense of the welfare of gas consumers and possibly smaller suppliers. In this case, the rule and assumptions of a typical Nash-Cournot equilibrium are not satisfied, and consumer gas prices are closer to monopoly prices despite the presence of several suppliers. A typical example of such a market and behavior is the presence of cartels in commodity markets.

Through non-parametric methods, an additional method for testing if European gas markets are integrated and if price convergence in the gas markets/ hubs has really occurred over the years is provided.

Overall, the results presented in this study lead to the conclusion that the institutional changes do not deliver on the objective of increasing competition among suppliers directly in a given hub (i.e., do not have the power to influence the number of competitors). It is more beneficial to concentrate on enhancing market integration, thus improving access to gas from the lowest-cost gas sources. Accordingly, the risk of price manipulation is reduced by ensuring that the different gas hubs in Europe are highly integrated.

The gas market in the Dutch TTF is anticipating an important decision on the speed of further reduction of domestic production amid uncertainties about the Groningen field production rates, which is the main provider of gas to customers located in Northwestern Europe. The security of supply is again at stake, a factor that increases Europe’s dependence on gas imports [135]. Therefore, diversification of supply and market integration is of primary importance in Europe.

Answer to Challenge 2: Due to the recent liberalization process, it is evident, that the gas markets in Europe are increasingly integrated and that long term oil indexed contracts are no longer the primary choice of customers. This is manifested by the gas suppliers' behavior in both gas markets that lean towards a competitive conduct that relies on maximizing profits, and by spot gas contracts that are increasingly being traded and trusted by the different stakeholder of the gas industry.

Challenge 3 – Main drivers for European wholesale Gas prices dynamics

Four different econometric methods were presented in chapter 6 to estimate parameters of regression: least square, maximum likelihood, machine learning gradient decent and least square optimization. These methods are used to compute the coefficients of a multivariate causal regression analysis, by linking them to a "use cause" in the subject of commodity pricing, such as natural gas.

Although there are several supply and demand fundamentals that govern the price formula, rational mathematical and economic interpretation presented contribute to the understanding of what causes natural gas price volatility in the German gas market in first place, and also succeeds in accurate price short term forecasts in a second place.

The performance comparison shows that there are some clear advantages in using machine learning gradient descent or least square optimization methods over the least square and maximum likelihood methods, such as the network's output matches the desired target with low mean absolute percentage error. In addition, the coefficients computed using the latter method is the most efficient and most accurate when used to forecast short term natural gas spot day ahead prices.

The results prove that, the non-parametric methods perform best in function approximation and this is line with the quality of the data that have low level of noise; in addition, the forecasted prices can be trusted and more accurate as the non-parametric algorithm makes no initial assumption about the data.

Answer to Challenge 3: The supply and demand variables (substitute fuel such as coal and oil, weather, exchange rates and storage capacity) chosen to forecast the wholesale German gas prices, gave accurate results. The advantage of the non-parametric econometric model and the calibration tests used (PCA and Gamma), have contributed in accurately forecasting short term gas prices. Both conclusions lead us to say that gas prices can be forecasted in such markets.

Challenge 4 – The role of Gas regulators: Assessing the need for further market intervention

Four econometric and mathematical methods are used collectively to estimate the level of information contained in gas prices in two separate wholesale gas markets, i.e. the European and the U.S. gas markets. The theories employed are Cournot Theory, Records Theory, Shannon Entropy, and Information Theory.

By analyzing the efficiency of the gas market and assessing the need for additional measures and intervention, the work of gas regulators in terms of market oversight is likely to be improved. The value of the information is based on three market signals: The possibility of non-competitive behavior by gas firms, market stability, and uncertainty in prices.

The findings suggest that the U.S. gas market is stable. The information value contained in the wholesale gas prices gives a clear indication that there is no need for additional market oversight by its regulator. However, this is not the case in the UK (the most developed European gas market), where results show signs of market instability and non-competitive behavior. In other words, some firms are not only focused on profit maximization; therefore, the wholesale prices are not solely the product of classical law of supply and demand.

Interestingly, the value of the additional information brought about by the signal analysis, included in both markets has contributed to reducing uncertainty. This makes the information carried in the gas prices of both markets powerful and efficient, although in asymmetric terms. The regulators in both markets can therefore use the two-step approach in order to assess the level of competition, price stability, price predictability, and act accordingly.

The originality of the two-step approach applied in this document can be summarized as follows: It is the first time that several multidisciplinary econometric methods have been combined to create a probabilistic structure assessing the underlying information of a gas market. Furthermore, the approach deals with a case where the information is neither completely absent nor perfectly known, which has rarely been dealt with in literature.

Answer to Challenge 4: *On the contrary to the U.S. market, a*ppropriate measures and actions need to be taken by the regulator in the U.K. gas market, in order to ensure that under no circumstances, the competing gas suppliers communicate and have the agreed-upon understanding to raise prices and profit margins at the expense of consumer welfare. Barrier to entry for new companies should also be considered and reduced by regulators in such markets, to increase competition and diversify supplies. Besides, the latter market show signs of price volatility and uncertainty in the medium term, and the long term. In such a case also, the regulator should anticipate such results by acting in favor of a continuous supply and demand equilibrium, by trying to diversify supply (indigenous production, imports, and storage), while also ensuring that the consumers have the appropriate infrastructure and financial means to buy such a commodity.

8.2 Strengths and weaknesses of each challenge/ model

A summary of the strengths and weaknesses of the Econometric models used in this thesis is described in Table 44, Table 45, Table 46 and Table 47. Each Table is followed by a detailed explanation of each of the points.

Challenge 1

Econometric models	Strength	Weakness
Records theory - Econometric models that focus on the tail of Distribution	-The theory of records is applied for the first time to gas markets -Records theory does not impose constraints on the quality and distribution of residuals - Record models beyond $i.i.d$ context are easily manipulated -The record theory findings and results are exact and non-asymptotic - The findings suggest that out of the three main regional gas markets, the Asian market seems to be less stable.	- Other Record models could have been used -Validation of the model results with classical in-sample, and out of sample method.

Table 44. Summary and synthesis of the Econometric models used challenge 1 – Nominal "one price" for gas markets.

__Advantages:__

- The theory of records is applied for the first time to gas markets, with the aim of testing the stability of three different regional gas markets. This solid measure of market stability would certainly be of interest to the gas markets that show signs of instability, such as those found in the Asian market. This is really important, especially knowing that the Yang model has succeeded in estimating the records for the Asia Pacific gas prices, and could be used to forecast the probability of future records.

- Standard linear models, stochastic volatility models, and regime-switching approaches are parametric methods, which are constrained by assumptions. However, the Records theory does not impose constraints on the quality and distribution of residuals which alleviates the use of multiple statistical tests that make the classical econometric approaches defined on hypothesis quasi-impossible to be entirely verified.

- Non-parametric, non-linear models such as machine learning could have been used, mainly because they are “constraint free” and therefore there is no need for additional tests (i.e. normality test for residuals, autocorrelation, etc.). However the records theory deals with extreme value of extreme values, therefore the number of available observations is generally small. Then, it is not reasonable to use machine learning models in particularly due to their need to have lots of training data in order to perform well.

- Record models beyond $i.i.d$ context are easily manipulated. Unlike other Extreme Value Theory models (Not to mention classical models).

- The record theory findings and results are exact and non-asymptotic.

<u>Weaknesses:</u>

- Other record models such as Linear Drift Model and Increasing Variance Model could have been used to compare between the models, however the models listed in chapter 4 (in its current format) is deemed extensive in my opinion.

- Statistical tests of a model's forecast performance are commonly conducted by splitting a given data set into an in-sample period, used for the initial parameter estimation and model selection, and an out-of-sample period, used to evaluate forecasting performance. However, a major reason of not considering the splitting principle is that I am dealing with records observation that are few in numbers, and if split, will influence the performance of the different considered record models and make them biased. Besides and in an attempt to validate the model results, I compare between theoretical and empirical results by considering the average number of records over all the study period.

Challenge 2

Econometric models	Strength	Weakness
Games Theory	-The non-parametric Cournot model has never been applied to test gas market concentration -The model bypasses statistical assumptions about underlying data, such as but not limited to cost and demand functions -Additional test for gas market integration is investigated - Results show that the new institutional changes introduced under the EU directives, have contributed to a better integration between the different gas hubs across Europe - The model results validates the fact that suppliers in the Dutch and Austrian markets follow the Nash-cournot equilibrium, even though one market is more liquid than the other.	- The assumption of a simple linear regression model used in the parametric method contributes to a low correlation coefficient -The cournot acceptance rate gives accurate results and detects which companies are competing under cournot assumptions. However when the suppliers are not abiding by cournot assumptions, this could imply many things: collusive behavior, strategic objectives that could be political, etc.

Table 45. Summary and synthesis of the econometric models used challenge 2 – European Gas price mechanism

Advantages:

- The models used in previous studies found in the literature are of parametric nature, which focuses on the use of empirical data and assumptions such as but not limited to cost and demand functions that yields a classical optimization problem in the context of non-linear programming. The data on cost and gas contracts are normally inaccessible to the public because of non-disclosure clauses, especially with regard to old legacy gas contracts. Additionally, most of the parametric analyses rely on statistical assumptions about underlying data. The results and conclusions can only be validated if the assumptions are correct. Therefore the use of a test that does not rely on such assumptions in the context of gas market suppliers' concentration and strategies is of added value to the scientific community.

- As explained in chapter 5, the non-parametric test allows to bypass many assumptions related to cost and demand functions. This study appears to be the first to customize the non-parametric theory and apply it to the natural gas market in an attempt to overcome the constraints of parametric models

- Additionally, by using models other than time series analysis, previously found in literature, the gas market integration in Europe is investigated.

- The model results validates the fact that suppliers in the Dutch and Austrian markets follow the Nash-cournot equilibrium and that they are rationally act to maximize their payoffs.

<u>Weaknesses:</u>

- The assumption of a simple linear regression model results (Used in the parametric method, in chapter 5) in a low correlation coefficient.

- The Cournot acceptance rate gives accurate results and detects which gas suppliers are competing under Cournot assumptions. However when the suppliers are not abiding by Cournot assumptions, this could imply many things: Collusive behavior, strategic objectives that could be political, etc. Further investigation and theoretical models should be envisaged in this regard to further investigate additional strategies.

Challenge 3

Econometric models	Strength	Weakness
Parametric and non -parametric regression methods of least square, maximum likelihood, machine learning gradient decent and least square optimization Econometric models that focus on the whole distribution	- This is the first time that both parametric and non-parametric methods are used for the same data set with the aim understand what drives gas prices in a first place and to forecast accurately short term gas prices - The advantage of the non-parametric econometric model and the calibration tests used (PCA and Gamma), have contributed in accurately forecasting short term gas prices -In-sample and out-of-sample were conducted for all methods – Model validation.	- Some of the essential assumptions of the MLR and VECM are rejected. Such as the normality and non-autocorrelation of residuals. This would force future modellers to rely on it with caution - Not all machine learning models were successful in forecasting accurate gas prices. In fact some models perform well in approximation, but poor performance when generalizing on a new set of data.

Table 46. Summary and synthesis of the Econometric models used challenge 3 – Main drivers for European wholesale Gas prices dynamics

Advantages:

- Analyzing the German gas hub spot prices data for supply and demand variables, the methods of least square, maximum likelihood, machine learning gradient decent and least square optimization are used to compute the coefficients of a multivariate causal regression analysis. The aim is to accurately forecast sort term gas prices. There are four different estimation techniques that are used in this study, divided into parametric and non-parametric estimation techniques. This mixture of methods allow the modeler to directly compare the performance of each.

- Since I am faced with a model that is composed of many observations, having more parameters than statistically needed is a possible outcome, therefore my model will be over-parameterized and over fitted. To bypass this problem, statistical tests such as PCA and Gamma are used.

  Two tests are introduced to simplify the data set, reduce the number of observations, and avoid over fitting because the data set is considered large. The first test is the Principal Component Analysis, PCA, which uses the orthogonal linear transformation that transforms the data to a new coordinate system, representing the new principal components. The aim is to get rid and omit the axis that holds the least amount of information, thus reducing the number of variables.

Artificial Neural Network is a non-parametric method that mostly use noon-linear activation functions in order predict and forecast commodity prices, and there is one vital question that remain to be answered. How many observations/ data points are needed to calibrate and forecast with accuracy? The second test that will help us answer this question is the use of the Gamma test to decrease the number of observations and data points for each variable.

The advantage of the non-parametric econometric model and the calibration tests used (PCA and Gamma), have contributed in accurately forecasting short term gas prices

- Statistical tests of a model's forecast performance were conducted by splitting a given data set into an in-sample period, used for the initial parameter estimation and model selection, and an out-of-sample period, used to evaluate forecasting performance.

Weaknesses:

- The higher the R^2, the better the linear model fits the data. However, in the first application, the Multi Linear regression model cannot be considered as the best fit for the simple fact that the assumption of normality and non-autocorrelation of residuals are not valid. Similarly the results of the Vector autoregressive analysis fail to meet initial model assumptions, such as the normality assumption; also, at least one of the inverses of roots is equal to one.

- Not all machine learning models were successful in forecasting accurate gas prices. In fact some models perform well in approximation, but poor performance when generalizing on a new set of data. Consequently, the models that perform well in approximation use estimation techniques with strong capabilities of memorizing but poor performance when generalizing on a new set of data. Unlike the two previous forecasting methods that are parametric and constrained by several assumptions on the data, the gradient descent and least squares methods are "constraint-free".

Challenge 4

Econometric models	Strength	Weakness
Information Theory (Probability Theory)	- The originality of the two-step approach applied in chapter 7 - The multidisciplinary econometric models have been combined to create a probabilistic structure, which is helpful in assessing the performance of the gas markets -This assessment is helpful to all stakeholders that participate in the gas value chain, most importantly, the regulators who play an important role in market oversight -Unlike the U.S. market, the U.K gas market, show signs of market instability and non-competitive behavior. This could mean that the wholesale prices are not solely the product of classical law of supply and demand.	-Signals chosen by the author might not be exhaustive -Subjectivity of the conditional probability that the market is either states (defined in chapter 7), after receiving the appropriate market signal.

Table 47. Summary and synthesis of the Econometric models used challenge 4 – The role of Gas regulators - Assessing the need for further market interventions

Advantages:

- The authors have identified three main metrics that can signal information in the hidden structure of the price values of both hubs. These metrics are based on econometric and mathematical methods, and are used to inform the regulator in each market about the following:

  Signal 1: Level of competition

  Signal 2: Market stability

  Signal 3: Volatility and uncertainty of prices

  In the first step approach, the authors have applied well-known mathematical theories in order to measure three signals. The signals and the means of measuring it with appropriate econometric methods is a completely new concept. Not to mention that the combination of such models was never used in commodity pricing (more specifically gas prices).

In the second approach, the level of information is measured. This is essential to evaluate the performance of the gas market and complements the results of the signal analysis (first approach).

These are assessed against the actions/ states that could be executed by the regulator of such markets. Two actions are identified; either to intervene in the market by taking legal actions, such as issuing new directives in order to ensure a stable supply and demand equilibrium and making sure that there are no abusive conduct by gas suppliers; or not to intervene.

The originality of the two-step approach applied in this document can be summarized as follows: It is the first time that several multidisciplinary econometric methods have been combined to create a probabilistic structure assessing the underlying information of a gas market. Furthermore, the approach deals with a case where the information is neither completely absent nor perfectly known, which has rarely been dealt with in literature.

Weaknesses:

- The choice of the conditional probability that the market is in state s_i after receiving the market signal is subjective. The probabilities p_{ik} are categorized into three classes: Low, Medium, and High. The assigned probabilities of 0.1, 0.5 and 0.9 are all possible events on the probability set.

8.3 Econometric models perspective and outlook on future of gas markets

Challenge 1

The analysis in chapter 4 used three record models to explain the different sets of data, each set consisting of gas prices observed in different gas markets. Pursuing further research and considering other record models for future research, such as Linear Drift Model and Increasing Variance Model, is needed.

Besides, record values were not computed in this analysis. In the future, researchers could study the joint distribution of record values based on a Markov chain analysis, which can provide more information to the inference procedure adopted in this document. Using this information, it is possible to predict future record values with a given confidence interval.

Interestingly, recent oil prices fell to the lowest in more than 17 years. Looking at it from a Records point of view, this is certainly a negative record (i.e. sudden drop in price). The decline in prices is due to several factors: Economic war between Saudi Arabia and Russia over oil supply and prices, lower seasonal demands due to the COVID pandemic, oversupply of oil, unavailability of additional storage capacities, etc. The oil prices falling trend will surely impact the gas markets that have a price mechanism indexed to oil. Among the three regional gas markets, this will surely be reflected in the Asian market where the gas/ LNG prices are mainly indexed to such a commodity.

One of the key factors that led to the drastic fall in oil prices recently is the lack of storage capacities that are essential to absorb the excess of supply. This is interesting from the European perspective. The fuel, is used to generate electricity and to heat, has already had a problematic year after a mild winter across the European continent. According to the EIA, European natural gas storage inventories as of March 1, 2020 were 60% full, the highest it has ever been in such times [136]. In addition, due to the recent COVID pandemic the summer demand is expected to be lower than average, this is a clear indication that there will little space if not any, available for storage capacities. Prices on the continent might drop and we will witness price volatility.

Challenge 2

Antitrust regulators should rely on further research and analyses of suppliers' behavior, other than the one that can be dedicated by the algorithm developed in chapter 5. In their quest to ban abusive behavior and in supervising mergers and acquisitions in wholesale and retail gas markets, future modelers should try to investigate and develop new methods that can detect collusive behavior.

They can also check if the strategies adopted by a supplier/company in a one- or few-stage game can be sustained or if it will lead to a different outcome. Besides, future research may focus on updating and enhancing the parameters used in this study, mainly with regard to the cost functions of the relevant big gas suppliers.

I would like to highlight two additional recent market developments that are happening in Europe and that could affect the functioning of the market as a whole in the coming future:

Netherlands and UK are two main gas producer in Europe. (Not counting Norway, since it is outside the EU). The Netherlands (represented by the TTF gas hub) is the most significant and mature hub, characterized by high number of active market participants, traded products, traded volumes, churn ratios, etc.

The gas market in the Dutch TTF is anticipating an important decision on the speed of further reduction of domestic production amid uncertainties about the Groningen field production rates, which is the main provider of gas to customers located in Northwestern Europe. The security of supply is at stake, a factor that increases Europe's dependence on gas imports. This will impact the TTF gas hub, especially in terms of liquidity, main element of price stability.

Additionally, the departure of the UK from the EU at the moment will also have an impact. EU energy law and all the EU energy directives are still in place until the end of 2020, more specifically 01 January 2021. This will have implications for the way gas is traded with the 27 EU member states. More specifically, access rules, network codes, capacity allocation mechanisms, prices are currently being revisited by regulatory authorities on both sides.

These are two examples that could shape the future of gas hubs in Europe.

Challenge 3

Although I was able to accurately forecast gas prices using appropriate econometric models in chapter 6, however, it is believed that additional mathematical parametric and non-parametric analysis should be done to further analyze and model natural gas prices.

Linear models can be employed with the condition of assuming Geometric Brownian Motion/ Radom walk instead of considering only white noise residuals; in addition, asymmetric GARCH as well as other non-linear econometric models can be employed as alternative to the "Gaussian" function while trying to model natural gas prices using machine learning methods.

Challenge 4

In chapter 7, the authors have chosen three market signals and four different econometric methods. It is believed that additional mathematical/statistical analysis can be used for this topic.

For further research, one can work on estimating the entropy generated (the third signal) using another discretization procedure. This work is a growing research track and needs a large number of observations. Besides, one can also work on creating estimators for the underlying probability distribution of each signal.

There is significant horizontal integration between the oil and gas industry as almost all major oil companies are also active in the natural gas business. Fall in oil prices below the brake even cost of most (if not all) unconventional oil producers in the U.S. (mainly unconventional producers) will certainly affect gas prices as well. Thus, if the upstream industry is not earning substantial revenues due to low oil prices, they are not endowed with the financial resources to invest in the exploration of natural gas fields. This will certainly lead to less supplies in the medium/long-term.

If the U.S. gas hubs loses one its main assets (i.e. liquidity) due to the slowdown in production, then this will have an enormous effect on prices and market development. How will this affect the Henry Hub gas prices?

References

[1] BP, BP Energy outlook, 2017. https://www.bp.com/content/dam/bp/pdf/energy-economics/energy-outlook-2017/bp-energy-outlook-2017.pdf.

[2] IEA, Outlook for Natural Gas, 2017. http://www.iea.org/publications/freepublications/publication/WEO2017Excerpt_Outlook_for_Natural_Gas.pdf.

[3] CEDIGAZ, Medium and Lon term Natural Gas outlook 2016, 2016.

[4] GECF, Global Gas Outlook, 2017.

[5] International Gas Union, World LNG Report, International Gas Union World Gas LNG Report. (2016) 88. http://www.igu.org/publications.

[6] EIA, Gas strategies LNG data service, 2015.

[7] International Energy Agency, Developing a natural gas trading hub in Asia: obstacles and opportunities, Paris, 2013.

[8] Xunpeng Shi; Hari M.P. Variam, Key elements for functioning gas hubs: A case study of East Asia, Natural Gas Industry B. 5 (2018) 167–176. doi:https://doi.org/10.1016/j.ngib.2018.03.001.

[9] Platts, No Title, (2017). https://www.platts.com/commodity/natural-gas.

[10] P. Economist, World LNG Database, (2015).

[11] T.N.C. Tom Kompas, A structural and stochastic optimal model for projections of LNG imports and exports in Asia-Pacific, Heliyon. 2 (2016).

[12] R. Davoust, Gouvernance europeene et geopolitique de l'energie, Bruxelles, 2008.

[13] D. Energy, Quarterly report on European Gas Markets, Bruxelles, 2016.

[14] H. Auer, Energy systems and grids, in: 2017.

[15] H. Hamie, A. Hoayek, H. Auer, Modeling the price dynamics of three different gas markets-records theory, Energy Strategy Reviews. 21 (2018) 121–129. doi:10.1016/j.esr.2018.05.003.

[16] A. Hoayek, H. Hamie, H. Auer, Modeling the Price Stability and Predictability of Post Liberalized Gas Markets Using the Theory of Information, Energies. 13 (2020). doi:https://doi.org/10.3390/en13113012.

[17] IGU, Wholesale Gas Price Survey, 2014.

[18] European parliament, The impact of the oil price on EU energy prices, 2014.

[19] E. Schultz, J. Swieringa, Price discovery in European natural gas markets, Energy Policy. 61 (2013) 628–634. doi:10.1016/j.enpol.2013.06.080.

[20] P. Heather, The evolution of European traded gas hubs, The Oxford Institute for Energy Studies. (2015).

[21] J. Stern, International gas pricing in Europe and Asia: A crisis of fundamentals, Energy Policy. 64 (2014) 43–48. doi:10.1016/j.enpol.2013.05.127.

[22] H. Hamie, A. Hoayek, H. Auer, M. Kamel, Northwestern European wholesale natural gas prices Comparison of several parametric and non-parametric forecasting methods, International Journal of Global Energy Issues. (2020). doi:10.1504/IJGEI.2020.10029052.

[23] A. Neumann, Linking natural gas markets - Is LNG doing its job?, Energy Journal. 30 (2009) 187–200. doi:10.2139/ssrn.1425519.

[24] P. Beatrice, The cost of price de-linkages between European gas hubs, The Oxford Institute for Energy Studies. (2015).

[25] International Gas Union (IGU), Wholesale Gas Price Formation-A global review of drivers and regional trends, 2015.

[26] Economic research institute for ASEAN and East Asia, Recommendations for a better functioning LNG market in Asia, The Institute of Electrical Engineers of Japan (IEEJ). (2015).

[27] B. Siliverstovs, G. L'Hégaret, A. Neumann, C. von Hirschhausen, International market integration for natural gas? A cointegration analysis of prices in Europe, North America and Japan, Energy Economics. 27 (2005) 603–615. doi:10.1016/j.eneco.2005.03.002.

[28] D. Zhang, Q. Ji, Further evidence on the debate of oil-gas price decoupling: A long memory approach, Energy Policy. 113 (2018) 68–75. doi:10.1016/j.enpol.2017.10.046.

[29] P. Heather, European traded gas hubs: An updated analysis on liquidity, maturity and barriers to market integration, The Oxford Institute for Energy Studies. (2017).

[30] B. Petrovich, The cost of price de-linkages between European gas hubs, The Oxford Institute for Energy Studies2. (2015).

[31] B. Petrovich, How strong is price correlation, 2013.

[32] D. KEMA, Country factsheets, 2013.

[33] ACER, Annual report on the results of monitoring the internal gas markets in 2016, 2017.

[34] D. Hulshof, J.P. van der Maat, M. Mulder, Market fundamentals, competition and natural-gas prices, Energy Policy. 94 (2016) 480–491. doi:10.1016/j.enpol.2015.12.016.

[35] J. Osička, L. Lehotský, V. Zapletalová, F. Černoch, B. Dančák, Natural gas market integration in the Visegrad 4 region: An example to follow or to avoid?, Energy Policy. 112 (2018) 184–197. doi:10.1016/j.enpol.2017.10.018.

[36] R. Egging, S.A. Gabriel, F. Holz, J. Zhuang, A complementarity model for the European natural gas market, Energy Policy. 36 (2008) 2385–2414. doi:10.1016/j.enpol.2008.01.044.

[37] T. Jansen, A. van Lier, A. van Witteloostuijn, T. Boon von Ochssée, A modified Cournot model of the natural gas market in the European Union: Mixed-motives delegation in a politicized environment, Energy Policy. 41 (2012) 280–285. doi:10.1016/j.enpol.2011.10.047.

[38] R.G. Egging, S.A. Gabriel, Examining market power in the European natural gas market, Energy Policy. 34 (2006) 2762–2778. doi:10.1016/j.enpol.2005.04.018.

[39] S.A. Gabriel, S. Kiet, J. Zhuang, A Mixed Complementarity-Based Equilibrium Model of Natural Gas Markets, Operations Research. 53 (2005) 799–818. doi:10.1287/opre.1040.0199.

[40] S.A. Gabriel, J. Zhuang, S. Kiet, A large-scale linear complementarity model of the North American natural gas market, Energy Economics. 27 (2005) 639–665. doi:10.1016/j.eneco.2005.03.007.

[41] F. Holz, C. von Hirschhausen, C. Kemfert, A strategic model of European gas supply (GASMOD), Energy Economics. 30 (2008) 766–788. doi:10.1016/j.eneco.2007.01.018.

[42] ACER, Framework guidelines on gas balancing in transmission systems, 2011.

[43] C. Miriello, M. Polo, The development of gas hubs in Europe, Energy Policy. 84 (2015) 177–190. doi:10.1016/j.enpol.2015.05.003.

[44] F. Asche, B. Misund, M. Sikveland, The relationship between spot and contract gas prices in Europe, Energy Economics. 38 (2013) 212–217. doi:10.1016/j.eneco.2013.02.010.

[45] T.G. Andersen, T. Bollerslev, F.X. Diebold, H. Ebens, The distribution of realized stock return volatility, Journal of Financial Economics. 61 (2001) 43–76. doi:10.1016/S0304-405X(01)00055-1.

[46] X. Mu, Weather, storage, and natural gas price dynamics: Fundamentals and volatility, Energy Economics. 29 (2007) 46–63. doi:10.1016/J.ENECO.2006.04.003.

[47] M. Giulietti, L. Grossi, M. Waterson, A Rough Analysis: Valuing Gas Storage, The Energy Journal. 33 (2012) 119–141. http://www.jstor.org/stable/23268107.

[48] TimeraEnergy, A framework for understanding European gas hub pricing, (2013).

[49] Maria Flouri et al., How does a natural gas supply interruption affect the EU gas security? A Monte Carlo simulation, Renewable and Sustainable Energy Reviews. 44 (2015) 785–796.

[50] L. Franza, Long-Term gas import contracts in Europe. The evolution in pricing mechanisms, Netherland, 2014. http://www.clingendaelenergy.com/inc/upload/files/Ciep_paper_2014-08_web_1.pdf.

[51] Fulwood Mike, Asian LNG trading Hubs: Myth or reality, 2018. https://energypolicy.columbia.edu/research/report/asian-lng-trading-hubs-myth-or-reality#_edn27.

[52] B. Petrovich, Do we have aligned and reliable gas exchange prices in Europe?, The Oxford Institute for Energy Studies. (2016).

[53] Z. Yang, R. Zhang, Z. Zhang, An exploration of a strategic competition model for the European Union natural gas market, Energy Economics. 57 (2016) 236–242. doi:10.1016/j.eneco.2016.05.008.

[54] A.J. McNeil, R. Frey, Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach, Journal of Empirical Finance. 7 (2000) 271–300. doi:10.1016/S0927-5398(00)00012-8.

[55] H.N.E. Byström, Managing extreme risks in tranquil and volatile markets using conditional extreme value theory, International Review of Financial Analysis. 13 (2004) 133–152. doi:10.1016/j.irfa.2004.02.003.

[56] Qinlu Chen & David E. Giles, Risk analysis for three precious metals: an application of extreme value theory, Econometrics Working Papers. (2014).

[57] K. Fong Chan, P. Gray, Using extreme value theory to measure value-at-risk for daily electricity spot prices, International Journal of Forecasting. 22 (2006) 283–300. doi:10.1016/j.ijforecast.2005.10.002.

[58] R. Herrera, A. Rodriguez, G. Pino, Modeling and forecasting extreme commodity prices: A Markov-Switching based extreme value model, Energy Economics. 63 (2017) 129–143. doi:10.1016/j.eneco.2017.01.012.

[59] Q. Liu, I. Wong, Y. An, J. Zhang, Asymmetric information and volatility forecasting in commodity futures markets, Pacific Basin Finance Journal. 26 (2014) 79–97. doi:10.1016/j.pacfin.2013.10.007.

[60] C. Aloui, R. Jammazi, The effects of crude oil shocks on stock market shifts behaviour: A regime switching approach, Energy Economics. 31 (2009) 789–799. doi:10.1016/j.eneco.2009.03.009.

[61] C. Aloui, S. Mabrouk, Value-at-risk estimations of energy commodities via long-memory, asymmetry and fat-tailed GARCH models, Energy Policy. 38 (2010) 2326–2339. doi:10.1016/j.enpol.2009.12.020.

[62] W. Chkili, S. Hammoudeh, D.K. Nguyen, Volatility forecasting and risk management for commodity markets in the presence of asymmetry and long memory, Energy Economics. 41 (2014) 1–18. doi:10.1016/j.eneco.2013.10.011.

[63] R. Aloui, M.S. Ben Aïssa, D.K. Nguyen, Conditional dependence structure between oil prices and exchange rates: A copula-GARCH approach, Journal of International Money and Finance. 32 (2013) 719–738. doi:10.1016/j.jimonfin.2012.06.006.

[64] E. Lindström, F. Regland, Modeling extreme dependence between European electricity markets, Energy Economics. 34 (2012) 899–904. doi:10.1016/j.eneco.2012.04.006.

[65] V. Marimoutou, B. Raggad, A. Trabelsi, Extreme Value Theory and Value at Risk: Application to oil market, Energy Economics. 31 (2009) 519–530. doi:10.1016/j.eneco.2009.02.005.

[66] H.N. Arnold, B. C., Balakrishnan, N. and Nagaraja, Records, John Wiley & Sons, New York, 1998.

[67] Nevzorov and Valery B., Records-mathematical theory, American Mathematical Society, Rhode Island, 2001.

[68] Rocco Ballerini and Sidney I. Resnick, Embedding sequences of successive maxima in extremal processes, with applications, Journal of Applied Probability. 24 (1987) 827–837.

[69] T. Robinson, Have European gas prices converged?, Energy Policy. 35 (2007) 2347–2351. doi:10.1016/j.enpol.2006.08.005.

[70] A. Neumann, A. Cullmann, What's the story with natural gas markets in Europe? Empirical evidence from spot trade data, in: 9th International Conference on the European Energy Market, EEM 12, 2012. doi:10.1109/EEM.2012.6254679.

[71] R. Golombek, E. Gjelsvik, K.E. Rosendahl, Effects of liberalizing the natural gas markets in Western Europe, Energy Journal. 16 (1995) 85–111. doi:10.2307/41322587.

[72] F.R. Aune, R. Golombek, S.A.C. Kittelsen, K.E. Rosendahl, Liberalizing the energy markets of Western Europe - A computable equilibrium model approach, Applied Economics. 36 (2004) 2137–2149. doi:10.1080/00036840310001641742.

[73] F. Forges, E. Minelli, Afriat's theorem for general budget sets, Journal of Economic Theory. 144 (2009) 135–145. doi:10.1016/j.jet.2008.03.002.

[74] I. Ray, L. Zhou, Game theory via revealed preferences, Games and Economic Behavior. 37 (2001) 415–424. doi:10.1006/game.2001.0845.

[75] Y. Sprumont, On the Testable Implications of Collective Choice Theories, Journal of Economic Theory. 93 (2000) 205–232. doi:10.1006/jeth.2000.2657.

[76] A. Carvajal, R. Deb, J. Fenske, J. Quah, Revealed Preference Tests of the Cournot Model, Econometrica. 81 (2013) 2351–2379. https://econpapers.repec.org/RePEc:ecm:emetrp:v:81:y:2013:i:6:p:2351-2379.

[77] B.T. Ewing, F. Malik, O. Ozfidan, Volatility transmission in the oil and natural gas markets, Energy Economics. 24 (2002) 525–538. doi:10.1016/S0140-9883(02)00060-9.

[78] O. Efimova, A. Serletis, Energy markets volatility modelling using GARCH, Energy Economics. 43 (2014) 264–273. doi:10.1016/j.eneco.2014.02.018.

[79] S. Nick, S. Thoenes, What drives natural gas prices? - A structural VAR approach, Energy Economics. 45 (2014) 517–527. doi:10.1016/j.eneco.2014.08.010.

[80] W.M. Fong, K.H. See, A Markov switching model of the conditional volatility of crude oil futures prices, Energy Economics. 24 (2002) 71–95. doi:10.1016/S0140-9883(01)00087-1.

[81] B. Lin, P.K. Wesseh, What causes price volatility and regime shifts in the natural gas market, Energy. 55 (2013) 553–563. doi:10.1016/j.energy.2013.03.082.

[82] T. Panagiotidis, E. Rutledge, Oil and gas markets in the UK: Evidence from a cointegrating approach, Energy Economics. 29 (2007) 329–347. doi:10.1016/j.eneco.2006.10.013.

[83] F. Fallahi, Causal relationship between energy consumption (EC) and GDP: A Markov-switching (MS) causality, Energy. 36 (2011) 4165–4170. doi:10.1016/j.energy.2011.04.027.

[84] F. Harrell, Regression Modeling Strategies, Springer, 2015.

[85] Gerda C; Nils H, Model Selection and Model Averaging, Cambridge Series in Statistical and Probabilistic Mathematics, 2008.

[86] S. Busse, P. Helmholz, M. Weinmann, Forecasting day ahead spot price movements of natural gas - An analysis of potential influence factors on basis of a NARX neural network, in: Multikonferenz Wirtschaftsinformatik 2012 - Tagungsband Der MKWI 2012, 2012.

[87] A. Khotanzad, H. Elragal, T.L. Lu, Combination of artificial neural-network forecasters for prediction of natural gas consumption, IEEE Transactions on Neural Networks. 11 (2000) 464–473. doi:10.1109/72.839015.

[88] N. Salehnia, M.A. Falahi, A. Seifi, M.H. Mahdavi Adeli, Forecasting natural gas spot prices with nonlinear modeling using Gamma test analysis, Journal of Natural Gas Science and Engineering. 14 (2013) 238–249. doi:10.1016/j.jngse.2013.07.002.

[89] F. Benedetto, G. Giunta, L. Mastroeni, On the predictability of energy commodity markets by an entropy-based computational method, Energy Economics. 54 (2016) 302–312. doi:10.1016/j.eneco.2015.12.009.

[90] A.F. Bariviera, L. Zunino, O.A. Rosso, Crude oil market and geopolitical events: An analysis based on information-theory-based quantifiers, Fuzzy Economic Review. 21 (2016) 41–51.

[91] W. Mensi, M. Beljid, S. Managi, Structural breaks and the time-varying levels of weak-form efficiency in crude oil markets: Evidence from the Hurst exponent and Shannon entropy methods, International Economics. 140 (2014) 89–106. doi:10.1016/j.inteco.2014.10.001.

[92] C.E. Shannon, Communication Theory of Secrecy systems, Bell System Technical Journal. 28 (1949).

[93] J. Yang, INFORMATION THEORETIC APPROACHES IN ECONOMICS, Journal of Economic Surveys. 32 (2018) 940–960. doi:10.1111/joes.12226.

[94] J. Crémer, A simple proof of Blackwell's "comparison of experiments" theorem, Journal of Economic Theory. 27 (1982) 439–443. doi:10.1016/0022-0531(82)90040-0.

[95] H.F. et al. Bohnenblust, Reconnaissance in Game Theory, RAND Memorandum. (1949). https://www.rand.org/pubs/research_memoranda/RM208.html.

[96] R. and O.J. Jones, Flexibility and Uncertainty, The Review of Economic Studies. 51 (1984) 13–32. doi:10.2307/2297702.

[97] A.S. Hoayek, G.R. Ducharme, Z. Khraibani, Distribution-free inference in record series, Extremes. 20 (2017) 585–603. doi:10.1007/s10687-017-0283-7.

[98] M.C.K. Yang, On the Distribution of the Inter-Record Times in an Increasing Population, Journal of Applied Probability. 12 (1975) 148–154. doi:10.2307/3212417.

[99] Nevzorov, Records for nonidentically distributed random variables, in: Proceedings of Fifth Vilnius Conference on Probability and Statistics, 1990: pp. 227–233.

[100] Karl Pearson, X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science. 50 (1990).

[101] A.W. H.B. Mann, On the choice of the number of class intervals in the application of the chi square teste, The Annals of Mathematical Statistics. (1942) 306–317.

[102] T. Inglot, T. Ledwina, Asymptotic optimality of data-driven Neyman's tests for uniformity, Annals of Statistics. 24 (1996) 1982–2019. doi:10.1214/aos/1069362306.

[103] J.C.W. Rayner, D.J. Best, The Choice of Class Probabilities and Number of Classes for the Simple $\chi ^{2}$ Goodness of Fit Test, The Indian Journal of Statistics. 44 (1982) 28–38. http://www.jstor.org/stable/25052234.

[104] F. C. Drost, A Limit Theorem for Some Modified Chi-Square Statistics when the Number of Classes Increases, Mathematical Statistics and Probability Theory. (1987) 43–58.

[105] Y.M. Bishop, S.E. Fienberg, P.W. Holland, Discrete Multivariate Analysis Theory and Practice, 2013. doi:10.1017/CBO9781107415324.004.

[106] J.C.W.R. Glen D. Rayner, Power of the Neyman smooth tests for the uniform distribution, Journal of Applied Mathematics and Decision Sciences. 5 (2001) 181–191.

[107] S.N. Majumdar, R.M. Ziff, Universal record statistics of random walks and lévy flights, Physical Review Letters. 101 (2008). doi:10.1103/PhysRevLett.101.050601.

[108] X. Shi, H.M.P. Variam, East Asia's gas-market failure and distinctive economics—A case study of low

oil prices, Applied Energy. 195 (2017) 800–809. doi:10.1016/j.apenergy.2017.03.091.

[109] T. Kompas, T.N. Che, Economic analysis of cost competitiveness of liquified natural gas exporters to the Asia-Pacific region, 2015.

[110] S. Thomas, Gas Regulation, Geeting Thre Deal through. (2018). https://gettingthedealthrough.com/area/15/jurisdiction/25/gas-regulation-austria/.

[111] IEA, Energy Policies of IEA countries, Netherland, 2014.

[112] G.H. Kuper, M. Mulder, Cross-border constraints, institutional changes and integration of the Dutch-German gas market, Energy Economics. 53 (2016) 182–192. doi:10.1016/j.eneco.2014.09.009.

[113] H. Anouk, The Dutch Gas Market: Trials, tribulations and trends, 2017.

[114] R. Golombek, E. Gjelsvik, K.E. Rosendahl, Increased competition on the supply side of the western European natural gas market, Energy Journal. 19 (1998) 1–18. doi:10.5547/ISSN0195-6574-EJ-Vol19-No3-1.

[115] C. Blitzer, Western European natural gas trade model, 1986.

[116] Wood Mackenzie, EU charges Gzprom over anticompetitive behavior, 2015.

[117] W. Sven-Olaf, Stoll; Klaus, A spot price model for natural gas considering temperature as an exogenous factor and applications, Journal of Energy Markets. 3 (2010) 113–128. doi:10.21314/JEM.2010.046.

[118] S. Johansen, Statistical analysis of cointegration vectors, Jounral of Economics Dynamics and Control. (1988).

[119] Y. Jiang, H. Guo, Y. Jia, Y. Cao, C. Hu, Principal component analysis and hierarchical cluster analyses of arsenic groundwater geochemistry in the Hetao basin, Inner Mongolia, Chemie Der Erde - Geochemistry. 75 (2015) 197–205. doi:10.1016/j.chemer.2014.12.002.

[120] Y.S. Park, G. Egilmez, M. Kucukvar, A Novel Life Cycle-based Principal Component Analysis Framework for Eco-efficiency Analysis: Case of the United States Manufacturing and Transportation Nexus, Journal of Cleaner Production. 92 (2015) 327–342. doi:10.1016/j.jclepro.2014.12.057.

[121] R. Remesan, M.A. Shamim, D. Han, Model data selection using gamma test for daily solar radiation estimation, Hydrological Processes. 22 (2008) 4301–4309. doi:10.1002/hyp.7044.

[122] S.E. Kemp, Gamma test analysis tools for non-linear time series, Department of Computing & Mathematical Sciences Faculty of Advanced Technology University of Glamorgan, 2006. http://www.r-project.org.

[123] A.J.J. Dafydd Evans, A proof of the Gamma test, Proceedings of the Royal Society A - Mathematical, Physical and Engineering Sciences. 458 (2002). doi:10.1098/rspa.2002.1010.

[124] M.T. Hagan, H.B. Demuth, M.H. Beale, Neural Network Design, Boston Massachusetts PWS. 2 (1995) 734. doi:10.1007/1-84628-303-5.

[125] EIA, Perspectives on the Development of LNG Market Hubs in the Asia Pacific Region, Washington, 2017. https://www.eia.gov/analysis/studies/lng/asia/pdf/lngasia.pdf.

[126] FERC, Market oversight - Federal Energy Regulatory Commission (FERC), FERC Website. (2016).

https://www.ferc.gov/market-oversight/guide/guide.asp?csrt=11553422067829331379.

[127] R.P.S. James, Gas Regulation - United States, Getting the Deal through. (2018). https://gettingthedealthrough.com/area/15/jurisdiction/23/gas-regulation-united-states/#link-28.

[128] X. Shi, Development of Europe's gas hubs: Implications for East Asia, Natural Gas Industry B. 3 (2016) 357–366. doi:10.1016/j.ngib.2016.11.001.

[129] A. et al. Brown, Gas Regulation - United Kingdom, Getting the Deal throughM. (2018). https://gettingthedealthrough.com/area/15/jurisdiction/22/gas-regulation-united-kingdom/.

[130] M.P. Haris, J. Tao, Role of governance in creating a commodity hub: A comparative analysis, Natural Gas Industry B. 3 (2016) 367–376. doi:10.1016/j.ngib.2016.12.001.

[131] W.H. Kruskal, W.A. Wallis, Use of Ranks in One-Criterion Variance Analysis, Journal of the American Statistical Association. 47 (1952) 583–621. doi:10.1080/01621459.1952.10483441.

[132] T.M. Cover, J.A. Thomas, Elements of Information Theory. Second Edition, 2005. doi:10.1002/047174882X.

[133] FERC, Characteristics of U.S. Natural Gas Transactions, 2016. https://www.cornerstone.com/Publications/Reports/Characteristics-of-US-Natural-Gas-Transactions-2016.

[134] R.M. Gray, Entropy and information theory, 2011. doi:10.1007/978-1-4419-7970-4.

[135] Wook-Mackenzie, Shell and ExxonMonil's Groningen future lies on shaky ground, 2018.

[136] EIA, European natural gas storage inventories are at record-high levels at the end of winter, (2020). https://www.eia.gov/todayinenergy/detail.php?id=43235 (accessed June 8, 2020).

[137] N. Salehnia, M.A. Falahi, A. Seifi, M.H. Mahdavi Adeli, Forecasting natural gas spot prices with nonlinear modeling using Gamma test analysis, Journal of Natural Gas Science and Engineering. 14 (2013). doi:10.1016/j.jngse.2013.07.002.

[138] OFGEM, Gas trading volumes and monthly churn ratio by platform (GB), (2018). https://www.ofgem.gov.uk/data-portal/gas-trading-volumes-and-monthly-churn-ratio-platform-gb.

Appendix to Chapter 4

Maximum Likelihood estimation method

Maximizing the Log-Likelihood function

$$log\, L(\gamma) = log\, \mathbb{P}[\delta_1, \dots, \delta_T; \gamma],$$

$$= N_T\, log(1 - 1/\gamma) + (T - N_T)\, log\, (1/\gamma) - log(1 - 1/\gamma^T) - \sum_{t=2}^{T} \delta_t\, log(1 - 1/\gamma^{t-1})$$

Thus, we must find the value $\hat{\gamma}$ of γ, which denotes our estimator by the maximum likelihood method, such that:

$$(d\, log\, L(\gamma)/d\gamma)_{\gamma=\hat{\gamma}} = 0 \tag{12}$$

The asymptotic behavior of this estimator is:

$$(\hat{\gamma} - \gamma)/\sqrt{I_T^{-1}(\gamma)} \to N(0{,}1), \tag{13}$$

Where $I_T(\gamma)$ denotes Fisher information. Note that the behavior of this estimator is also distribution-free. The importance of calculating the asymptotic behavior is double folded: It shows that the behavior of the estimator is distribution-free, which means independent from the choice of the underlying distribution Y. In addition, it gives the possibility to construct the confidence intervals of the parameter γ for a given confidence level alpha.

Goodness-of-fit for a Yang model

In the following section N_T denotes the number of records after the warm-up period and T the present time.

On the basis of a partition (of disjoint sub-sets) $\Pi_1 \cup \ldots \cup \Pi_K$ of the set $\{1,2,\ldots,\infty\}$, n_k, $1 \le k \le K$ denotes the number of Δ_{L_n} which fall within Π_k with $n_1 + \cdots + n_K = m - 1$ (Because for $N_T = m$ records it corresponds $m - 1$ inter-records). In addition, I denote $\pi_k(\gamma) = \sum_{j \in \Pi_k} p_j(\gamma)$, $1 \le k \le K$. The Pearson chi-square statistic is given by:

$$\chi(\gamma) = \sum_{k=1}^{K} \left(n_k - (m-1)\pi_k(\gamma)\right)^2 / ((m-1)\pi_k(\gamma)) \tag{14}$$

The value of this statistic is compared to $x^2_{K-1,1-\alpha}$ the quantile of order $(1-\alpha)$ of the chi-square with $K - 1$ degrees of freedom, denoted χ^2_{K-1}. When $\chi(\gamma) > x^2_{K-1,1-\alpha}$, the test rejects the hypothesis H_0 at the confidence asymptotic level α. Thus, the observed values of Δ_{L_n} are not in agreement with the geometric distribution and the model does not fits Yang.

However, the statistic $\chi(\gamma)$ is unusable because the parameter γ is unknown. I must therefore estimate it. To do this, I calculate the value $\tilde{\gamma}$ that minimizes $\chi(\gamma)$. Then, the usable statistic is

$$\chi(\tilde{\gamma}) = \underset{\gamma}{argmin}\, \chi(\gamma). \tag{15}$$

According to a classical result [105], if the data comes from a geometric distribution, the statistic $\chi(\tilde{\gamma}) \to \chi^2_{K-2}$ converges in distribution to a chi-square with $K - 2$ degrees of freedom. If $\chi(\tilde{\gamma})$ exceeds $x^2_{K-2,1-\alpha}$ (the quantile of order $1 - \alpha$ of the chi-square with $K - 2$ degrees of freedom), the test rejects, at the confidence level α, the null hypothesis that the Δ_{L_n} follow a Geometric distribution, which casts doubt on Yang's model.

List of Figures

Figure 1. Major trade movements 2018 - Trade flows worldwide, billion cubic meters 15
Figure 2. Main regional Natural gas markets (in Green) 16
Figure 3. Monthly Gas prices in three different Natural gas markets 17
Figure 4. Global spatial distribution of proved gas reserves relative to Europe 21
Figure 5. Research questions and challenges 24
Figure 6. European wholesale market 32
Figure 7. Three different types of commodity forecasting 41
Figure 8. Monthly gas imports to Austria and indigenous production 61
Figure 9. Monthly gas imports The Netherlands and indigenous production 62
Figure 10. Monthly Gas Prices for the two Gas Hubs, Euros/MWh 63
Figure 11. Payoff matrix 64
Figure 12. Frequency distribution of gas prices 77
Figure 13. Bi-plot of the principal components factors (F1 and F2) 81
Figure 14. Observations vs. Gamma & Standard Error 84
Figure 15. Back propagation neural network structure illustrative drawing 85
Figure 16. Plots of gas price forecasts using the Neural Network Gradient Decent method 93
Figure 17. Plots of gas price forecasts using least square optimization method 94
Figure 18. The variation of the values predicted by all models, from the observed values 96
Figure 19. Indigenous production and monthly gas imports to U.S. 98
Figure 20. Indigenous production and monthly gas imports to UK 99
Figure 21. Monthly Gas Prices of both Gas markets, USD/MMBtu 103

List of Tables

Table 1. Indexation level of European gas contracts 27

Table 2. Notation for the models applied in chapter 4 45

Table 3. Data collection - Records Theory 47

Table 4. Number of records and record index 52

Table 5. Goodness of fit test results at a confidence level of 5% 53

Table 6. Pearson chi-squared test results at a confidence level of 5% 53

Table 7. Probability of Records for each market 54

Table 8. Results of empirical and theoretical probabilities 55

Table 9. Notation for the models applied in chapter 5 57

Table 10. VTP participation 59

Table 11. Churn ratio 60

Table 12. Cost assumptions parameters 65

Table 13. Correlation coefficient and significance 68

Table 14. Strategic output of each supplier - Austria - 2013 69

Table 15. Parametric method - yearly gas prices results (in USD per Tons of Oil Equivalent) 69

Table 16. Actual values - yearly gas prices results (in USD per Tons of Oil Equivalent) 70

Table 17. Cournot acceptance rate 70

Table 18. Notation for the parametric methods applied in chapter 6 72

Table 19. Notation for non-parametric methods applied in chapter 6 73

Table 20. Data collection - Econometric models 75

Table 21. Descriptive statistics 76

Table 22. Likelihood-ratio test results 79

Table 23. List of exogenous and endogenous variables 79

Table 24. Correlation matrix 80

Table 25. Rotated factor loadings of principal components on variables 80

Table 26. Least square results 90

Table 27. Maximum likelihood results 91

Table 28. Parameter restrictions on the lagged relationships using the VECM model for time (t-1) at a confidence level of 5 % 92

Table 29.Mean absolute percentage error (MAPE) of the different models 95

Table 30. Notation for the models applied in chapter 7 97
Table 31. U.S. and UK traded volumes and churn ratio.................. 99
Table 32.Data collection 102
Table 33. Information structure conditional probabilities. 107
Table 34. Goodness of fit test results 109
Table 35. Number of records and record index.................. 110
Table 36. Probability of Records results 110
Table 37. Entropy results 111
Table 38. Synthesis of signal results for the EU market 112
Table 39. Synthesis of signal results for the U.S. market 113
Table 40. Information structure conditional probabilities for the EU market 115
Table 41. Information structure conditional probabilities for the U.S. market 116
Table 42. Conditional entropy of each signal 117
Table 43. "Posterior" and "apriori" entropies 117
Table 44. Summary and synthesis of the Econometric models used challenge 1 – Nominal "one price" for gas markets. 122
Table 45. Summary and synthesis of the econometric models used challenge 2 – European Gas price mechanism 124
Table 46. Summary and synthesis of the Econometric models used challenge 3 – Main drivers for European wholesale Gas prices dynamics 126
Table 47. Summary and synthesis of the Econometric models used challenge 4 – The role of Gas regulators - Assessing the need for further market interventions 128