NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS
DEPARTMENT OF ECONOMICS

M.Sc. in BUSINESS ADMINISTRATION, ANALYTICS AND INFORMATION SYSTEMS
(Π.Μ.Σ. Διοίκηση, Αναλυτική Και Πληροφοριακά Συστήματα Επιχειρήσεων)

THESIS

Forecasting Natural Gas Prices in the West European Market: A Comparative Analysis of Forecasting Models

RICHARD CONSTANDINE ALLAN

SUPERVISOR
DR. KYRIAKOS DRIVAS

ATHENS
2023

NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS

DEPARTMENT OF ECONOMICS

M.Sc. in BUSINESS ADMINISTRATION, ANALYTICS AND INFORMATION SYSTEMS
(Π.Μ.Σ. Διοίκηση, Αναλυτική Και Πληροφοριακά Συστήματα Επιχειρήσεων)

THESIS

Forecasting Natural Gas Prices in the West European Market: A Comparative Analysis of Forecasting Models

RICHARD CONSTANDINE ALLAN

SUPERVISOR
DR. KYRIAKOS DRIVAS

ATHENS
2023

I hereby certify the submitted thesis and the work presented is personal and that all sources and materials used have been properly referenced in the text and bibliography.

Richard Constandine Allan

Dedication

To my parents Marditsa and Graham, my brother Robert, my partner Marialena, and to the memory of my grandparents, Kostas, Maria, Jessy, and Jack.

This thesis is dedicated to all of you who have helped me become the person I am today.

Thank you!

Acknowledgements

I would like to thank my parents and brother for their unwavering support throughout my academic journey. Their encouragement and belief in me have been a constant source of motivation, and I am forever grateful for everything they have done for me.

Additionally, I would like to thank my partner, Marialena. She provided unwavering emotional support, understanding, and patience throughout the thesis process. I am truly blessed to share my life with her.

I also want to thank my closest friends, Andreas and Kostis, for their support and understanding. They were always there to listen to me when I needed to vent, and their help with research and feedback on my work was invaluable.

Furthermore, I would like to thank my professor, Dr. Kyriakos Drivas for his guidance and feedback on my work. His expertise and dedication to my success were instrumental in shaping my thesis and my academic journey as a whole.

Finally, I would like to express my deep gratitude to my head professor, Dr. Ioannis C. Demetriou, and the National and Kapodistrian University of Athens. I am truly blessed to have studied in the most prestigious Greek institution under the tutoring of such a gifted and dedicated professor.

I am truly honored to have had the opportunity to pursue my studies, and I am deeply grateful to everyone who has played a role in my academic journey.

Table of Contents

List of Tables 4

List of Figures 4

Abstract 5

Περίληψη 6

1. Introduction 7

2. Definitions 9

2.1 Natural Gas 9

2.2 Trading Hubs 9

2.3 Dutch Title Transfer Facility (TTF) 10

2.4 Brent Crude Oil 10

2.5 Coal (API2) CIF ARA Continuous Contract 11

2.6 Forecasting Models 11

3. Literature Review 12

3.1 Approaches 12

3.2 Variables 13

4. Aim of Study 15

4.1 Research Hypotheses 15

5. Methodology 16

5.1 Models 16

5.1.1 Multiple Linear Regression 16

5.1.2 Seasonal Auto-Regressive Integrated Moving Average with eXogenous factors 17

5.1.3 Recurrent Neural Network 18

5.1.4 Support Vector Regression 20

5.2 Data 21

5.2.1 Natural Gas Price 22

5.2.2 Crude Oil Price 22

5.2.3 Coal Price 23

5.2.4 Storage Capacity 23

5.2.5 Average Temperature of Western Europe 24

5.3 Tool 24

5.4 Process 25

5.4.1 Exploratory Data Analysis 25

5.4.2 Dataset Splitting 28

5.4.3 Supervised Learning 29

6. Results 34

6.1 Descriptive Statistics 34

6.2 Correlation and Autocorrelation Analysis 36

6.3 Model Evaluation 38

7. Conclusions 40

8. Discussion 41

9. Limitations 43

10. Suggestions for Future Research 44

References 45

List of Tables

Table 1 Variable Description ... 22
Table 2 Time Series Matrix ... 26
Table 3 Formulas of Evaluation Metrics ... 30
Table 4 SARIMAX Results ... 31
Table 5 Measures of Central Tendency ... 34
Table 6 Measures of Variability ... 35
Table 7 Measures of Shape ... 35
Table 8 Correlation Matrix ... 36
Table 9 Evaluation Metrics ... 39

List of Figures

Figure 1 TTF Price Time Series ... 27
Figure 2 Capacity Time Series ... 27
Figure 3 Coal Price Time Series ... 27
Figure 4 Brent Price Time Series ... 27
Figure 5 Temperature Time Series ... 27
Figure 6 Correlation Heatmap ... 37
Figure 7 Autocorrelation Plot (ACF) ... 37
Figure 8 Partial Autocorrelation Plot (PACF) ... 38

Abstract

The use of natural gas has increased significantly in recent years, with many countries shifting towards this energy source to meet their growing energy demands while reducing their carbon emissions. West Europe is particularly significant, given the region's high demand for natural gas and its reliance on imports. The complexity of the European energy market, combined with the volatility of natural gas prices, presents unique challenges for forecasting models. The drivers that impact energy prices and the best models for forecasting prices are important subjects of research and analysis.

The aim of this study was to conduct a comparative analysis of four forecasting models, including Multiple Linear Regression (MLR), Recurrent Neural Network (RNN) with Long-Short Term Memory (LSTM), Seasonal Auto-Regressive Integrated Moving Average with eXogenous factors (SARIMAX), and Support Vector Regression (SVR), to assess which model provides the most accurate predictions for natural gas prices in Western Europe. The Dutch TTF Spot Last Price was the variable acting as the price of natural gas in Western Europe, and the regressors included in the models were the available natural gas in European storage, the average temperature of Western Europe, and the European prices of coal and crude oil. The procedures of data processing, analysis, and forecasting were conducted with the use of "Python" libraries and tools.

The analysis revealed that the SVR model achieved the highest level of accuracy among the tested models. Notably, the findings indicated that complex machine learning models did not necessarily outperform traditional statistical models in forecasting natural gas prices. Further research is necessary to investigate the performance of different forecasting models and their sensitivity to natural gas price drivers.

Keywords*: Natural Gas · Forecasting Models · Machine Learning · Western Europe · SVR · MLR · SARIMAX · RNN · Dutch TTF Spot Price*

Περίληψη

Τα τελευταία χρόνια, η ζήτηση του φυσικού αερίου έχει αυξηθεί σημαντικά, με πολλές χώρες να στρέφονται προς αυτή την πηγή ενέργειας για να καλύψουν τις αυξανόμενες ενεργειακές τους ανάγκες, μειώνοντας παράλληλα τις εκπομπές διοξειδίου του άνθρακα. Η Δυτική Ευρώπη παρουσιάζει ιδιαίτερο ενδιαφέρον, δεδομένης της υψηλής ζήτησης για φυσικό αέριο και της εξάρτησής από τις εισαγωγές. Η πολυπλοκότητα της ευρωπαϊκής αγοράς ενέργειας, σε συνδυασμό με τη μεταβλητότητα των τιμών του φυσικού αερίου, δημιουργεί σημαντικές προκλήσεις για τα μοντέλα πρόβλεψης. Οι παράγοντες που επηρεάζουν τις τιμές της ενέργειας και η εύρεση του βέλτιστου μοντέλου πρόβλεψης τιμής, αποτελούν σημαντικά αντικείμενα έρευνας και ανάλυσης.

Σκοπός της παρούσας μελέτης ήταν η συγκριτική ανάλυση μεταξύ των μοντέλων Πολλαπλής Γραμμικής Παλινδρόμησης (MLR), Επαναλαμβανόμενου Νευρωνικού Δικτύου (RNN) με Μακροπρόθεσμη-Βραχυπρόθεσμη Μνήμη (LSTM), Εποχιακού Αυτοπαλινδρομικού Ολοκληρωμένου Κινητού Μέσου Όρου με Εξωγενείς παράγοντες (SARIMAX) και Διανυσματικής Παλινδρόμησης Υποστήριξης (SVR), με στόχο την ανάδειξη του ακριβέστερου μοντέλου πρόβλεψης τιμών φυσικού αερίου στη Δυτική Ευρώπη. Η τελευταία τιμή Dutch TTF Spot αποτέλεσε τη μεταβλητή που ενεργούσε ως τιμή του φυσικού αερίου στη Δυτική Ευρώπη και οι παλινδρομικοί παράγοντες που συμπεριλήφθηκαν στα μοντέλα ήταν το διαθέσιμο φυσικό αέριο στις ευρωπαϊκές αποθήκες, η μέση θερμοκρασία της Δυτικής Ευρώπης και οι ευρωπαϊκές τιμές του άνθρακα και του αργού πετρελαίου. Οι διαδικασίες επεξεργασίας, ανάλυσης και πρόβλεψης των δεδομένων πραγματοποιήθηκαν με τη χρήση βιβλιοθηκών και εργαλείων "Python".

Με βάση τα αποτελέσματα της ανάλυσης, το μοντέλο SVR πέτυχε το υψηλότερο επίπεδο ακρίβειας μεταξύ των συγκρινόμενων μοντέλων. Ειδικότερα, τα ευρήματα έδειξαν ότι τα σύνθετα μοντέλα μηχανικής μάθησης δεν υπερέχουν απαραίτητα των παραδοσιακών στατιστικών μοντέλων στην πρόβλεψη των τιμών του φυσικού αερίου. Απαιτείται περαιτέρω έρευνα για τη διερεύνηση των επιδόσεων των διαφόρων μοντέλων πρόβλεψης και της ευαισθησίας τους στους παράγοντες που επηρεάζουν τις τιμές του φυσικού αερίου.

Λέξεις-Κλειδιά: *Φυσικό Αέριο · Μοντέλα Πρόβλεψης · Μηχανική Μάθηση · Δυτική Ευρώπη · SVR · MLR · SARIMAX · RNN · Τιμή TTF Spot*

1. Introduction

Natural gas is the world's second most actively traded energy commodity, especially in the European Union (EU) (Čeperić et al., 2017; Hafner & Raimondi, 2022). Because of the substitution of coal and the increasing demand for cleaner-burning fuels, forecasting natural gas prices has become one of the most crucial initiatives that many companies in different industries have to undertake. Natural gas price forecasting accuracy is critical since these predictions are utilised for managing potential risk and reducing the difference between supply and demand. It is also essential for economic planning, energy investment, and environmental preservation, and it serves as a vital guide for the efficient execution of energy policy and planning. In order to improve the accuracy of future predictions, experts are continuing to extensively research and analyse natural gas price forecasting models.

This study focuses on natural gas pricing dynamics in Western Europe as the latter is the industrial epicentre of the European Union and one of the largest regional gas markets in the world. The aforementioned area is highly dependent on natural gas imports since it utilises natural gas as a means of heating, transportation, and energy production (Haifner & Raimondi, 2022). The different gas markets are also highly integrated with each other, creating a West European market, including countries such as the Netherlands, France, Germany, Belgium, Italy, Austria, Denmark, and the UK (Papież et al., 2022; Broadstock et al., 2020; Kuper & Mulder, 2016). These factors result in an unpredictable pricing environment during periods of economic crises and diplomatic conflicts (Kalogiannidis et al., 2022; Richter & Holz, 2015).

This study attempts, through the use of data analytics, energy economics, and machine learning, to assess the accuracy of different forecasting models. The models examined in this study include Multiple Linear Regression (MLR), Seasonal Auto-Regressive Integrated Moving Average with eXogenous factors (SARIMAX), Recurrent Neural Network (RNN) with Long-Short Term Memory (LSTM), and Support Vector Regression (SVR).

Title Transfer Facility (TTF) spot price is the variable of interest since it acts as a benchmark price for all the other European trading hubs. The variables relating to natural gas that were examined in this study, include the average temperature in Western Europe and the available natural gas in European storage. Additional regressors included the prices of energy commodities, those being crude oil, and coal. Historically, fluctuations in oil and coal prices tended to cause alterations to the price of natural gas (Nick & Thoenes, 2014). By analysing these factors, the models can make more accurate forecasts about future price movements.

This thesis intends to contribute to the academic literature by being a link between the different approaches to forecasting natural gas prices. As most researchers utilise either statistical or advanced machine learning models, this study will attempt to connect all their individual findings. This research can be utilised to assist policymakers, guide investment strategies, and optimize energy supply chains. The challenges faced during the process of research and analysis were, amongst others, the interdisciplinary nature of the subject, scarcity of available data, and developing the programming code to organise and analyse the data.

The structure of this study was the following:

- **Chapter 2** contains important definitions for understanding this study,
- **Chapter 3** analyses the existing literature on the various methodologies and approaches used in predicting future gas prices,
- **Chapter 4** states the aim of the study,
- **Chapter 5** explains the data and models used in this study, and the steps of data processing, analysis, and forecasting,
- **Chapter 6** contains the results of the analysis,
- **Chapter 7** presents the conclusions drawn from the results,
- **Chapter 8** promotes a discussion regarding the findings of the study,
- **Chapter 9** underlines the limitations of the study, and
- **Chapter 10** suggests proposals for similar research in the future.

2. Definitions

Given the technical nature of the study and to provide clarity and precision, this chapter defines key terminology used throughout the thesis.

2.1 Natural Gas

Natural gas is a fossil fuel that is predominantly made up of methane (CH_4), along with small traces of other hydrocarbons and non-hydrocarbon gases such as nitrogen and carbon dioxide (Atwater et al., 2023). It is a versatile source of energy that can be utilised for various applications, such as generating electricity, heating, cooking, and transportation (U.S. Energy Information Administration, 2022).

Natural gas has the benefit of being a relatively clean-burning fossil fuel, generating lower amounts of greenhouse gases and air pollutants than coal and oil (Atwater et al., 2023). Even so, natural gas production and transportation can cause environmental issues such as air and water pollution, habitat damage, and greenhouse gas emissions (U.S. Energy Information Administration, 2022; Atwater et al., 2023).

In general, it seems to be a reliable source of energy, an important driver for various industries and economies, and a tool for political stability and international relations. A fair system of production and trade of natural gas can promote peace, prosperity, and an alternative fuel source for reducing greenhouse gas emissions.

2.2 Trading Hubs

Natural gas trading hubs are critical components of the global energy infrastructure as they serve as exchange platforms for buyers and sellers to trade natural gas molecules at both spot and futures prices (Tong et al., 2014). These hubs are strategically located at the centre of gas infrastructure networks, including pipelines and LNG terminals. A hub serves as the network's principal price point for natural gas (Roinioti et al., 2014). Gas trading hubs can be both physical and virtual, depending on their use. Physical hubs are geographical locations that function as transit sites for natural gas transportation as well as storage facilities (Tong et al., 2014). Virtual hubs are trading platforms for financial transactions between larger parties and organisations (Honoré, 2013). The benefit of virtual hubs is the ability to trade any gas that has paid a price for network access, as opposed to physical hubs that only allow the trading of gas that is physically traveling through a certain place (Roinioti et al., 2014).

In 2022, there were over 12 active gas trading hubs in 10 different European countries (Chestney, 2022). In order to integrate its natural gas markets, the European Union advocated the construction of a number of virtual (regional) trading hubs that often overlapped national borders. The Dutch TTF (Title Transfer Facility) is considered to be the primary hub for natural gas trading, with the British NBP (National Balancing Point) following closely behind (Shi & Variam, 2018). Other important trading hubs include the German NCG (NetConnect Germany) and Gaspool, the Italian PSV (Punto di Scambio Virtuale), the Austrian VTP (Virtual Trading Point), the Belgian Zeebrugge Beach (ZEE) and ZTP, the Spanish PVB, the French Point Exchange Gaz (PEG) and the Czech VOB (Chestney, 2022).

2.3 Dutch Title Transfer Facility (TTF)

The Dutch TTF, or Title Transfer Facility, is a virtual hub located in the Netherlands (Hafner & Luciani, 2022). It was established in 2003 by Gasunie Transport Services (GTS) and it is a platform for the trading of natural gas contracts, including spot, forward, and futures contracts (Roinioti et al., 2014).

The TTF hub has established itself as the primary gas trading hub in continental Europe due to its strategic location and connection to major natural gas pipelines and storage facilities. As a result, it has become the benchmark for pricing natural gas contracts across the region, as highlighted by Jotanovic & D'Ecclesia (2021) and Papież et al. (2022).

The TTF hub offers increased price transparency and competition in the natural gas market since it enables the trading of natural gas between buyers and sellers through a number of physical distribution locations (Papież et al., 2022). The platform is operated by the Intercontinental Exchange (ICE) and the hub is overseen by the Netherlands Authority for Consumers and Markets (ACM), and the Dutch Ministry of Economic Affairs and Climate Policy.

2.4 Brent Crude Oil

Brent Crude Oil refers to the type of crude oil that is extracted from the North Sea and traded on the ICE. This type of “light” crude oil is named “Brent”, since it is extracted in the Brent Oil Field, in the East Shetland Basin. It is considered a major pricing benchmark for crude oil, since it serves as one of the three primary benchmark prices for purchasing oil globally, along with West Texas Intermediate (WTI) and Dubai/Oman. Essentially, Brent affects the price of oil in Europe, WTI is the dominant benchmark in the Americas, and Dubai/Oman influences the Asian market (Corbett, 1990; Kurt, 2022).

Brent crude oil is a mixture of hydrocarbons with different molecular weights and structures (C_nH_{2n+2}). This type of crude oil can be refined to produce various petroleum products, such as gasoline, butane, diesel, heating oil, propane, and jet fuel. It is well known for its low density, low viscosity, and low sulphuric content (Chen, 2022).

Even though Brent makes up two-thirds of the global crude oil trade, its position is heavily threatened by depleting supply from the North Sea and the general shift towards more sustainable sources of energy. These factors may impact its long-term viability as an energy source.

2.5 Coal (API2) CIF ARA Continuous Contract

The Coal (API2) CFI ARA Continuous Contract is a futures contract for API2 coal traded on the ICE. The price of this contract is affected by the API2 index, which is considered to be a benchmark price evaluation for coal imported into Northwest Europe (Mosquera-López & Nursimulu, 2019). The API2 index is calculated by Argus Media and McCloskey Group, two independent pricing agencies that specialise in commodity markets (Fernández Alvarez, 2022). These two agencies do not own the physical assets related to coal production and transportation. Instead, they jointly own the intellectual property rights of the index and provide pricing information for the transparent trade of coal.

As referenced by the “CIF” abbreviation, the price of the API2 index includes all costs that result from the trade of coal (Cost, Insurance, and Freight) (Liu & German, 2017). “ARA” stands for the ports of Amsterdam, Rotterdam, and Antwerp, which are the three major ports in NW Europe. These three ports collectively serve as the European trading hub of various commodities such as coal, metals, and oil (Fernández Alvarez, 2022). Finally, the term “continuous contract” refers to a futures contract that does not have a defined expiration date and will continue to be until one of the contract's parties ends it (Kagan, 2021).

Overall, the Coal (API2) CIF ARA Continuous Contract is a widely used futures contract for API2 coal, providing traders and companies with a tool for hedging and speculating on the price of this important energy commodity.

2.6 Forecasting Models

Forecasting models are tools that are used to predict future trends and outcomes based on past data and statistical analysis (Hyndman & Athanasopoulos, 2018). There are various types of forecasting models that can be used in different contexts. By selecting the appropriate forecasting model, researchers and analysts can better anticipate future trends and outcomes, and make more informed decisions (Hyndman & Athanasopoulos, 2018).

The most important widely used forecasting models are time series analysis, regression analysis, machine learning models (or AI models), judgmental forecasting, ensemble forecasting, econometric forecasting, and simulation forecasting (Wang & Chaovalitwongse, 2011).

When choosing a forecasting model, it is important to take into account the nature of the data, the purpose of the forecast, along with the strengths and weaknesses of each model.

3. Literature Review

This chapter will provide a comprehensive overview of the current state of knowledge on the topic of forecasting natural gas prices. It will also include the research hypotheses of this study.

3.1 Approaches

Given the intricate nature of predicting natural gas prices, numerous approaches have been suggested as potential solutions to address this challenge. The majority of studies have used individual or mixed statistical and machine learning models, in order to make accurate forecasts. As a general guideline, researchers typically compare at least two to five different models when conducting analysis for natural gas price forecasting. These models can either be statistical, machine learning, or other, depending on the researcher's expertise and available data.

Statistical models such as Autoregressive Integrated Moving Average (ARIMA) (Erdogdu, 2010; Hosseinipoor et al., 2016), Generalized Autoregressive Conditional Heteroskedasticity (GARCH) (Hosseinipoor et al., 2016; Lv & Shan, 2013) and Vector Autoregression (VAR) (Aminu, 2019; Nick & Thoenes, 2014), have been the most consistently used models when forecasting gas prices, due to their efficiency and simplicity. Researchers have also attempted to use more advanced methods such as the Auto-Regressive Integrated Moving Average with eXogenous factors (ARIMAX) model, in order to incorporate external variables into the analysis (Seo, 2021). Even though these models are a staple of traditional forecasting, they bear some limitations when it comes to capturing complex market dynamics and incorporating external factors that can influence natural gas prices.

In general, Machine Learning models such as Artificial Neural Networks (ANN) and Support Vector Regression Machines (SVM) will outperform statistical models when attempting to forecast natural gas prices (Azadeh et al., 2012; Salehnia et al., 2013; Seo, 2021). Lately, researchers have been utilising various machine learning models to predict natural gas prices and subsequently evaluate and compare the outcomes. For example, Su et al. (2019) compared the accuracy of models such as ANN, SVR, GPM, and GPR, and concluded that ANN was the most accurate out of the group. Other less common models will even outperform traditional machine learning algorithms. Ali (2020) applied the Linear-squares Boosting model (LSBoost) to outperform all four models utilised by Su et al. (2019). While machine learning models have shown promising results, there are cases where traditional statistical models prevail. An indicative paradigm can be viewed in the case of Jin & Kim (2015) where GARCH outperformed ANN in terms of forecasting accuracy. Another important example was the case of Čeperić et al. (2017) where traditional time series models, such as AR and ARIMA, dominated over NN and SVR.

In summary, while each of these techniques can be applicable for predicting natural gas prices, they all have limitations that should be taken into consideration. It is crucial to carefully consider the assumptions and performance of each technique and consider using a combination of said techniques. By combining statistical and machine learning models, the outcome would yield a more accurate and robust forecast by capturing the complexity and nuances of the data, while incorporating a wide range of factors that influence prices.

3.2 Variables

The determinants used in forecasting natural gas spot prices vary depending on the methodology used for a given analysis and time horizon. Most researchers have concluded that a handful of factors consistently affect the fluctuations of natural gas prices. A comprehensive analysis of the various determinants can provide valuable insights into natural gas prices' short- and long-term price trends.

A plethora of studies agree that the most important factor affecting natural gas prices is the fluctuation of crude oil prices. Oil is the most actively traded energy commodity in the world, and therefore has the ability to affect the prices of other energy commodities. Based on past and recent studies, crude oil spot (Brown & Yucel, 2008; Hartley & Medlock, 2014; Hartley et al., 2008; Nick & Thoenes 2014; Rubaszek & Uddin, 2020) and futures prices (Chen et al., 2023; Li & Song, 2023; Linn & Zhu, 2004; Mu, 2007; Prokopczuk et al., 2021) play a prominent role when forecasting natural gas prices. However, in recent years, more and more studies (Batten et al., 2017; Lin & Li, 2015; Wang et al., 2019) argue that crude oil and natural gas prices have decoupled. Specifically, for the TTF Day-Ahead price, Hulshof et al. (2016) determined that the price of Brent Oil affects the price by a minuscule percentage.

Temperature also seems to be a prominent determinant, as extreme temperatures affect inventory and final demand. Natural gas is a predominant heating source in North-West Europe, resulting in higher demand during the cold months of Autumn and Winter. Sudden changes in temperature and increased demand during colder months result in price spikes and fluctuation (Wang et al., 2019). Specifically for Western Europe, Nick & Thoenes (2014) found that abnormal temperatures affect natural gas prices in the short term. Finally, Hulshof et al. (2016) determined that the number of heating degree days is one of the most significant factors that affect TTF Day-Ahead prices.

Another significant determinant is gas storage and inventory. A number of studies have determined that an increase in gas storage levels will decrease the price of the commodity (Mu, 2007). Nick & Thoenes (2014) found that storage and supply shortfalls will affect natural gas prices in the short term, while Gay et al. (2009) and Halova et al. (2014) reported that supply and demand shocks impact natural gas prices during injection and withdrawal seasons (April-October and November-March, respectively). Finally, if natural gas storage levels fall below the expected point, the Day-Ahead price of TTF will increase (Hulshof et al., 2016).

Coal price is also a prominent determinant in several studies, given that both commodities can be used for heating and power generation. While the relationship between coal and natural gas can vary over time and by region, there appears to be a meaningful link between the two fuels that can impact energy markets. For example, Nick & Thoenes (2014) argued that coal prices affect the price of natural gas in the long run. Furthermore, Li et al., (2017) observed the strong influence of natural gas on coal prices and their integration in the European markets. Regarding the TTF Day-Ahead price, Hulshof et al. (2016) reported that while coal prices positively impact the commodity of interest, the effect is insignificant, potentially due to the competition between these two commodities in the energy market.

Other determinants worth mentioning are imports of liquid natural gas (Nick & Thoenes, 2014), heating oil prices (Hu et al., 2014), price of carbon credits (Hulshof et al., 2016), and natural gas supply and demand (Su et al., 2019).

4. Aim of Study

The aim of this study was to identify the most accurate model for forecasting natural gas prices and to evaluate the effectiveness of complex machine learning models in comparison to traditional statistical models. By achieving this aim, the study aimed to provide valuable insights into improving the accuracy of forecasting techniques for natural gas prices.

The variables and models used in this analysis were selected based on a thorough review of the existing literature on natural gas price forecasting. The TTF Spot Last Price was chosen as the dependent variable, as it represents the benchmark price of natural gas in Western Europe.

This study incorporated a range of independent variables that are known to impact the price of natural gas in Western Europe. Specifically, the average temperature of Western Europe and the amount of natural gas in European storage were included as internal factors because they directly influence the demand and supply of natural gas within the region. In addition to these internal factors, the prices of two energy commodities - Coal and Brent oil - were included as external factors. These factors impact natural gas prices indirectly through their influence on the energy sector supply and demand dynamics.

The models chosen for this study were based on two key factors: their ability to effectively incorporate multiple independent variables, and their overall complexity as a forecasting model. To this end, four different models were chosen to be tested and compared: Multiple Linear Regression (MLR) and Seasonal Auto-Regressive Integrated Moving Average with eXogenous factors (SARIMAX) were selected as traditional statistical models, while Support Vector Regression (SVR) and Recurrent Neural Network (RNN) with Long-Short Term Memory (LSTM) were selected as more complex machine learning models. These models belong to different categories of forecasting models, including Statistical (MLR), Time Series (SARIMAX), Support Vector Machines (SVR), and Neural Networks (RNN).

4.1 Research Hypotheses

Drawing on the analysis presented, the research hypotheses of this thesis are as follows:

- ***Hypothesis 1***: Complex machine learning models will exhibit superior performance in forecasting natural gas prices, as compared to traditional statistical models.

- ***Hypothesis 2***: RNN+LSTM, as a sophisticated machine learning technique, will outperform all other forecasting models in terms of its accuracy in predicting natural gas prices.

5. Methodology

This chapter provides an overview of the data and models used in the analysis, as well as a detailed account of the data cleaning, analysis, and forecasting processes.

5.1 Models

In order to ensure the forecast's accuracy, a total of four models were selected based on comparable research and with the purpose of conducting a comparative analysis of their individual results. The details of each of the aforementioned models which were utilised in the present research are illustrated below.

5.1.1 Multiple Linear Regression

Multiple linear regression (MLR) is a statistical modeling technique used to examine the relationship between a dependent variable and multiple independent variables. In other words, it is used to estimate how various factors affect a particular outcome variable (Schmidt & Finan, 2018).

The MLR equation is:

$$Y = \beta 0 + \beta 1X1 + \beta 2X2 + \ldots + BpXp + \varepsilon$$

Where:

- Y is the dependent variable,
- X1 to Xp are the independent variables,
- β0 is the intercept, and β1 to βp are the coefficients (also known as slopes) that represent the impact of each independent variable on the dependent variable, and
- ε is the error term, which represents the variability in the dependent variable that cannot be explained by the independent variables.

In general, MLR can be used to analyse the relationship between variables and to make predictions about the values of the dependent variable based on the values of the independent variables (Schmidt & Finan, 2018).

One of the strengths of MLR is that it can model the relationships between natural gas prices and multiple independent variables, such as supply and demand factors, which can allow for the identification of significant drivers of price movements (Gujarati & Porter, 2009). In addition, MLR can provide a quantitative estimate of the effect of each independent variable on natural gas prices, allowing analysts to better understand the underlying factors that affect prices (Field, 2013). Furthermore, MLR can be used to generate forecasts based on the values of the independent variables (Field, 2013).

However, MLR also has several limitations that should be carefully considered. For example, MLR assumes that the relationships between the independent variables and the dependent variable are linear, which may not always be the case (Tabachnick & Fidell, 2013). This can lead to biased or inaccurate forecasts, particularly if the relationships are nonlinear. In addition, MLR assumes that the relationships between the independent variables and the dependent variable are constant over time, which may not be the case for natural gas prices, as they can be affected by a range of dynamic factors (Tabachnick & Fidell, 2013). Furthermore, MLR is sensitive to outliers and influential data points, which can significantly affect the estimated coefficients and the model's performance (Gujarati & Porter, 2009). Finally, MLR assumes that the independent variables are independent of each other, which may not always be the case in practice. Multicollinearity between the independent variables can lead to unreliable estimates of the coefficients and inflated standard errors (Jeon, 2015).

In summary, while MLR is a powerful and versatile technique, it is important to keep its limitations in mind when using it to model data. By being aware of these limitations and taking steps to address them, MLR can be an effective tool for analysing and modeling complex data.

5.1.2 Seasonal Auto-Regressive Integrated Moving Average with eXogenous factors

SARIMAX, also known as Seasonal Auto-Regressive Integrated Moving Average with eXogenous factors, is a widely used statistical modeling technique for time series analysis and forecasting (Brockwell & Davis, 2002). It is an extension of the ARIMA (Autoregressive Integrated Moving Average) model, which is a popular tool for modeling and forecasting time series data that do not exhibit seasonal patterns. SARIMAX, on the other hand, is capable of capturing seasonal variations that are present in time series data (Hyndman & Athanasopoulos, 2018).

The equation of the SARIMAX model is the following:

$$y(t) = AR(p) + MA(q) + AR(P)m + MA(Q)m + e(t)$$

Where:

- y(t) is the time series at time t,
- AR(p) represents the autoregressive component of order p,
- MA(q) represents the moving average component of order q,
- AR(P)m represents the seasonal autoregressive component of order P with a period of m,
- MA(Q)m represents the seasonal moving average component of order Q with a period of m, and
- e(t) is the residual at time t.

A SARIMAX model comprises three key components: the autoregressive (AR) component, the integrated (I) component, and the moving average (MA) component. These components describe how the model should respond to the data and how it should adjust over time. The AR component models the relationship between an observation and a certain number of past observations, also known as lags, whereas the MA component models the relationship between an observation and a certain number of past errors. The I component accounts for non-stationarity in the data by differencing the series, which helps to stabilize the variance and make the series stationary (Brockwell & Davis, 2002).

SARIMAX models are capable of modeling seasonal variations by incorporating seasonal differencing, which is a similar process to the non-seasonal differencing used in ARIMA models. Seasonal differencing involves taking the difference between an observation and the corresponding observation from the previous season. This can be repeated for multiple seasons if necessary (Hyndman & Athanasopoulos, 2018).

SARIMAX models can also incorporate external variables that may affect the time series being analysed. Exogenous variables are external factors that are not part of the time series itself but may influence it. Examples of exogenous variables that could be used in a SARIMAX model for natural gas prices include economic indicators, such as GDP or inflation, or weather data, such as temperature or precipitation.

However, SARIMAX models do have some limitations. They are dependent on having a sufficient amount of data available for accurate modeling and forecasting, and they assume that the relationship between the dependent variable and the exogenous variables is linear. In addition, it can be challenging to identify the appropriate order of the AR, I, and MA components, as well as the appropriate order of the seasonal components, which can impact the accuracy of the model's forecasts (Shumway & Stoffer, 2011).

Overall, SARIMAX is a powerful tool for modeling and forecasting time series data, particularly when the data exhibit seasonal patterns and is influenced by exogenous variables. However, it is crucial to carefully evaluate the limitations and assumptions of the model when applying it to a specific problem.

5.1.3 Recurrent Neural Network

Recurrent neural network (RNN) is a type of neural network that is specifically designed for sequential data processing by using feedback connections t that enable them to maintain information over time. Unlike feedforward neural networks, which process input data in a single pass, RNNs can operate on input sequences of arbitrary length, making them well-suited for tasks such as speech recognition, natural language processing, and time series forecasting (Bengio et al., 1994; Goodfellow et al., 2016;).

The main feature of RNNs is to create a loop in the network architecture that enables information to be transmitted from one time step to the next. At each time step, the network takes an input vector x(t) and a hidden state vector h(t-1) as inputs and produces an output vector y(t) and a new hidden state vector h(t) as outputs.

The equations that govern the computation of these vectors are known as the forward equations of the RNN:

$$h(t) = f(W_x * x_t + W_h(t-1) + b)$$
$$y(t) = g(W_{yh(t)} + b_y)$$

Where at time t:

- x is the input,
- y is the output,
- h is the hidden state,
- f is the activation function,
- W_x and W_h are weight matrices for the input and hidden state,
- b is the bias term,
- g is the output activation function,
- W_y is the weight matrix for the output, and
- b_y is the output bias term.

The hidden state vector h(t) serves as a kind of "memory" for the network, allowing it to store information about previous time steps and use it to make predictions about future time steps (Bengio et al., 1994; Goodfellow et al., 2016).

The first equation computes the new hidden state vector h(t) by taking a weighted sum of the input vector x(t) and the previously hidden state vector h(t-1) and passing the result through an activation function f. This allows the network to "remember" information from previous time steps and use it to make predictions about future time steps.

The second equation computes the output vector y(t) by taking a weighted sum of the new hidden state vector h(t) and passing the result through another activation function g. This output vector can be used to make predictions about the task at hand, such as predicting the next value in a time series.

RNNs can be utilised for natural gas price forecasting. This can be achieved by “feeding” the model historical prices as inputs and training the network to predict future prices. The network can be trained using various optimization techniques such as Back Propagation Through Time (BPTT) or the Long Short-Term Memory (LSTM) algorithm. Once the network is trained, it can be used to generate forecasts for future prices based on the latest available data.

RNNs have several strengths, including their ability to capture temporal dependencies in data, their flexibility for a variety of tasks, and their computational efficiency due to weight sharing across time steps (Bianchi et al., 2017).

However, RNNs also have several weaknesses. One common problem is the vanishing gradient problem, where the gradients become very small during backpropagation, hindering the network's ability to capture long-term dependencies (Goodfellow et al., 2016). RNNs can also forget information from earlier time steps as they process new data, making it challenging to retain long-term memory. Finally, RNNs can be computationally expensive during training, particularly if the sequence length is long (Bianchi et al., 2017).

To mitigate some of these weaknesses, techniques such as LSTM and GRU have been developed. These techniques improve RNNs' ability to capture long-term dependencies and retain long-term memory (Goodfellow et al., 2016).

In summary, RNNs have several strengths for processing sequential data, but they also have limitations that need to be addressed through careful design and optimization (Bianchi et al., 2017).

5.1.4 Support Vector Regression

Support Vector Regression (SVR) is a type of supervised learning algorithm that is used for regression tasks. It is a variation of the Support Vector Machine (SVM) algorithm and is commonly used in machine learning applications to forecast future numerical values based on past data (Hastie et al., 2009).

The basic idea behind SVR is to find a hyperplane in a high-dimensional space that can best separate the data into different categories while minimizing the margin error. In regression, the aim is to find a hyperplane that best fits the data while minimizing the prediction error.
The formula for SVR can be expressed as follows:

$$y = wT * \Phi(x) + b$$

Where:

- y is the predicted value,
- x is the input data,
- w is the weight vector,
- b is the bias, and
- Φ(x) is a feature vector that maps the input x to a higher-dimensional space.

The objective of SVR is to minimize the error term while ensuring that the predicted values fall within a certain range (Abe, 2005). This is accomplished by:

- minimizing $(1/2)||w||\^2 + C\Sigma(\varepsilon i + \varepsilon i^*)$,
- subject to the constraints: yi - wTΦ(xi) - b <= εi wTΦ(xi) + b - yi <= εi*

Where:

- w is the weight vector,
- Φ(xi) is the feature vector,
- b is the bias term,
- εi and εi* are the slack variables, and

- C is the penalty parameter that controls the trade-off between the margin and the errors (Abe, 2005).

SVR can be used for forecasting gas prices by training the model on historical data that includes various factors that can influence the prices, such as weather, supply and demand, political events, and economic indicators. The trained model can then be used to predict future natural gas prices based on the current values of these factors (Hastie et al., 2009).

SVR is a powerful regression method with several strengths. One of its key strengths is its ability to handle non-linear relationships between variables. This is achieved by using non-linear kernel functions, which can map the input variables to a higher-dimensional space and separate them with a hyperplane (Geron, 2019). Another strength of SVR is its robustness to outliers. The method includes slack variables that allow for some errors beyond the margin, which can make it less sensitive to outliers than other regression methods (Muller & Guido, 2016). Finally, SVR is flexible in terms of the choice of the kernel function, which allows the user to select the kernel that best suits the particular prediction problem (Bishop, 2006).

Despite its many strengths, SVR has some weaknesses. One of its main weaknesses is its sensitivity to the choice of hyperparameters. The performance of SVR can be highly dependent on the selection of hyperparameters such as the regularization parameter and the kernel parameters (Geron, 2019). Tuning these hyperparameters can be time-consuming and requires careful selection. Another weakness of SVR is its computational intensity. The training of an SVR model can be computationally intensive, especially for large datasets and complex kernel functions (Muller & Guido, 2016). Finally, interpreting the results of an SVR model can be difficult. The model is defined in a high-dimensional space, and the weights assigned to the input variables in the model may not be easily interpretable (Murphy, 2012).

To summarise, SVR possesses numerous advantages that grant it the ability to be a highly effective regression method. However, certain drawbacks of this technique must not go unnoticed, while applying it for predictive purposes involving natural gas prices or other problems. Careful hyperparameter selection and model interpretation can help mitigate some of these weaknesses, but they remain important considerations when using SVR for forecasting natural gas prices or other real-world prediction problems.

5.2 Data

The datasets examined in this thesis consist of daily data from January 2nd, 2014 to April 29th, 2020. For each dataset, a brief description followed by the source provider, method of acquisition and/or calculation, and unit of measure is provided. Additional miscellaneous information about the datasets is also provided, when necessary. The variables' names, definitions, units of measure, and sources are summarised in Table 1.

Table 1 Variable Description

Variable	Description	Source	Unit of Measure
TTF Price	TTF Spot Last Price	European Energy Excange (EEX)	EUR/MWh
Brent Price	Europe Brent Oil Spot Last Price FOB	Energy Information Administration (EIA)	EUR/Barrel
Coal Price	Coal (API2) CIF ARA Continuous Contract Last Price	MarketWatch	EUR/Tonne
Capacity	Storage Capacity of European Natural Gas	Gas Infrastructure Europe (GIE)	Percentage (%)
Temperature	Average Temperature between 13 West European Countries	European Climate Assessment & Dataset	°C

5.2.1 Natural Gas Price

This dataset was provided by the European Energy Exchange (EEX) in Microsoft Excel Worksheet (XLS) format. After contacting the Data Source Department and signing an agreement of confidentiality, the dataset was the first to be collected for this study. This dataset included daily “Spot” and “Within-Day” prices for three virtual trading points of natural gas in Europe, those being Title Transfer Facility (TTF), National Balancing Point (NBP), and Gaspool. It also included columns such as traded volume, traded contracts, contract name, open, high, low, and last price of each trading point.

For this study, the Dutch TTF Last Price was used as the dependent variable due to its status as the standard reference price for natural gas trade across Europe. (Jotanovic & D'Ecclesia, 2021; Papież et al., 2022). After erasing the extra columns, the column was renamed “TTF Price”. The dataset was structured as a two-column matrix with 2312 rows. The “TTF Price” column contained the latest price at which natural gas was traded for immediate delivery. The daily price was measured in terms of euros per megawatt hour (EUR/MWh). It is important to note that the dataset contained several null values on random dates.

Source: https://www.eex.com/

5.2.2 Crude Oil Price

This dataset was downloaded from the website of the U.S. Energy Information Administration (E.I.A), in XLS format. A formal agreement was not required for the use of the data by this Organisation. This dataset included the daily spot price of European Brent Oil FOB. FOB stands for “Free On Board”, meaning that this commodity is charged at the actual price recorded at the port of the producing country. This price includes any applicable discounts, rebates, or premiums, and does not require adjustment for credit terms. Essentially, it is the price that was paid in full for the product being loaded onto a transport vessel (U.S. Energy Information Administration, 2016).

For this study, the European Brent Oil price was used as a determinant of the dependent variable, since it is considered the benchmark price of oil trade in Europe (Corbett, 1990) and an important driver of natural gas prices (Rubaszek & Uddin, 2020). The dataset was structured as a two-column matrix with 1631 rows. It had fewer rows than the dependent variable since it did not provide values of the traded commodity during the weekends. The column of the traded price was renamed "Brent Price" and was measured in terms of euros per barrel (EUR/Barrel). In the oil industry, an oil barrel contains 42 US gallons or 158.987 litres of crude oil (U.S. Energy Information Administration, 2016).

Source: https://www.eia.gov/

5.2.3 Coal Price

This dataset was downloaded in XLS format, from the MarketWatch website which is owned by Dow Jones & Company. A formal agreement wasn't required for the use of the data by this Organisation. The dataset included daily data on the last price of the Coal (API2) CIF ARA Continuous Contract. It also included the first, highest, and lowest price of the contract, as well as the price difference within a day of trading.

For this study, the last price of the contract was used as a determinant of the dependent variable, since it is considered a benchmark price of the coal trade in North-West Europe (Mosquera-López & Nursimulu, 2019), and its impact on natural gas price fluctuations (Hulshof et al., 2016). After deleting all the extra columns, the column was renamed "Coal Price". The dataset was structured as a two-column matrix with 1631 rows. It had fewer rows than the dependent variable since it did not provide values of the traded commodity during the weekends. Coal was measured in euros per metric tonne. A metric tonne is approximately 1000 kilograms or 2204.62 pounds.

Source: https://www.marketwatch.com/

5.2.4 Storage Capacity

This dataset was acquired from the website of the Gas Infrastructure Europe (GIE) Aggregated Gas Storage Inventory (AGSI), in XLS format. A formal agreement wasn't required for the use of the data by this Organisation. The dataset contained information on available gas in storage, storage capacity, trend, gas injection and withdrawal, consumption, and working gas volume.

For this study, the storage capacity was used as a determinant of the dependent variable, since storage shortfalls do affect prices (Nick & Thoenes, 2014) and especially the TTF price (Hulshof et al., 2016). After deleting all the extra columns, the column was renamed "Capacity". The dataset was structured as a two-column matrix with 2312 rows. Capacity was measured as the percentage of available gas in storage facilities.

Source: https://agsi.gie.eu/

5.2.5 Average Temperature of Western Europe

This dataset was generated by conducting personal calculations to determine the average temperature of Western Europe. To obtain the necessary data, 27 meteorological stations were selected from nine Western European countries, with each country contributing three stations. The countries included in this analysis were England, Finland, Germany, Luxemburg, the Netherlands, Scotland, Switzerland, Spain, and Sweden. The website of the European Climate Assessment & Dataset was able to provide text files (TXT) with daily measurements of the average temperature from every station. A formal agreement wasn't required for the use of the data by this Organisation. These files included data about station and source identifiers, mean temperature, and quality codes.

For this study, the average temperature was used as an independent variable, since cold weather and heating demand days do affect natural gas prices (Nick & Thoenes 2014; Wang et al., 2019) and specifically the TTF Day-Ahead price (Hulshof et al., 2016). The daily mean temperature was derived by taking the average of temperature measurements obtained from these stations. After deleting all the extra columns, the column was renamed "Temperature". The dataset was structured as a two-column matrix with 2312 rows. "Temperature" was measured in degrees Celsius (°C).

Source: https://www.ecad.eu/

5.3 Tool

The Python programming language, developed by Guido van Rossum (Avouris et al., 2020), was utilised for data cleaning and analysis in this study. Python is a widely used high-level language among data analysts and engineers due to its open-source nature and versatility in performing tasks such as data processing, cleaning, analysis, visualization, and machine learning.

Python works by utilising libraries and packages that provide a wide range of tools and functions for various programming tasks. These libraries were developed by an active community of programmers who contribute to the maintenance and development of the language.

The libraries utilised in this study included:

- Pandas for data manipulation and transformation,

- NumPy and SciPy (stats) for data analysis,
- Seaborn for data visualization,
- Matplotlib (pyplot) for the creation of plots and charts,
- Statsmodels (tsa.seasonal, api, graphics.tsaplots) for statistical modelling,
- Sklearn (metrics, svm, preprocesing, model_selection, linear_model), for forecasting,
- TensorFlow (keras.models, keras.layers) for building neural networks,
- Datetime for formatting dates, and
- Pmdarima for fitting data into ARIMA models.

These libraries were imported into Jupyter Lab, to create a clear and interactive data analysis notebook.

5.4 Process

Before the process of analysing the datasets could begin, each dataset was converted into a Comma-Separated Values file (CSV), to be read and processed by Python. Once the datasets were in a usable format, the process of data cleaning could begin.

5.4.1 Exploratory Data Analysis

Performing exploratory data analysis (EDA) is a crucial component of the data analysis process. The following paragraphs outline the step-by-step approach to conducting EDA.

Using the Pandas library, five distinct data frames were created by reading the CSV file of each variable. A data frame is a two-dimensional tabular data structure (Avouris et al., 2020). Similarly to spreadsheets or SQL tables, they are comprised of rows and columns on which data can be stored and processed. In Python, they are commonly used in data analysis by implementing the Pandas library (Avouris et al., 2020). The data frames were renamed into simple acronyms, to reflect their content (Brent, Coal, Cap, Temp, and TTF).

In order to process the data accurately, it was essential to have a clear understanding of the structure and format of the data frames. The "head" and "count" functions were used to display the first few rows of each data frame and the number of rows in each column, respectively. The "describe" function was also utilised to provide a more comprehensive view of each data frame, by summarising the key statistical properties of each column. After applying these functions, three data frames were found to have 2312 rows (TTF, Temp, and Cap) while the other two data frames contained 1631 rows each (Coal and Brent).

During the analysis, it was discovered that the "TTF Price" column had an extreme maximum value of 72.375, which is significantly higher than the other values in the column. Additionally, the "Temperature" column had an extreme minimum value of -999.90, which is much lower than the expected range of temperatures. The “count” function also revealed that the “TTF” data frame had 14 missing values.

To initiate the data analysis process, it was necessary to merge the five data frames into a single entity. This would provide a more comprehensive view of the data by combining information from separate tables. However, before merging the data frames, it was crucial to ensure that their shapes and values were correct and comparable. This involved checking for consistency in column size and data types.

The first step of the process included converting all data frames into time series. This was achieved by setting the date columns as the indexes for each time series. Additionally, all indexes were converted into “Date-Time” objects since they needed to have the same format before merging. The data type of the “Capacity” column was converted from percentage to decimal format, as some Python libraries cannot process percentages. The outlier in the “Temperature” column was replaced with the mean value, as it was considered a data entry error. On the other hand, the "TTF Price" outlier was not altered as it was a valid value in the time series and could be attributed to a sudden change in the market. Finally, the time series “Brent” and “Coal” needed to be the same size as all other time series. This was handled by resampling the time series and calculating the mean value for each missing day.

After completing all the previous steps, the time series were merged into a single entity. Any rows that contained missing values in the “TTF Price” column were removed from the merged time series. The resulting merged time series was structured as a five-column matrix, consisting of 2297 rows. Table 2 displays the first and last few observations of the merged dataset.

Table 2 Time Series Matrix

Date	TTF Price	Temperature	Capacity	Coal Price	Brent Price
02-01-2014	26.30	8.8	0.69	60.88	81.38
03-01-2014	26.03	9.1	0.69	59.94	80.35
04-01-2014	26.75	8.1	0.69	59.94	80.35
-	-	-	-	-	-
27-04-2020	5.92	13.8	0.62	34.29	29.64
28-04-2020	6.02	11.8	0.62	33.72	29.77
29-04-2020	5.87	12.7	0.62	34.20	28.70

Further observation of trends or patterns was achieved by visualizing each column. Line plots can visualise data over time while providing information such as upward and downward trends, seasonal fluctuations, and sudden changes in the data. The “matplot.pyplot” function was utilised for each of the five columns (Figures 1-5).

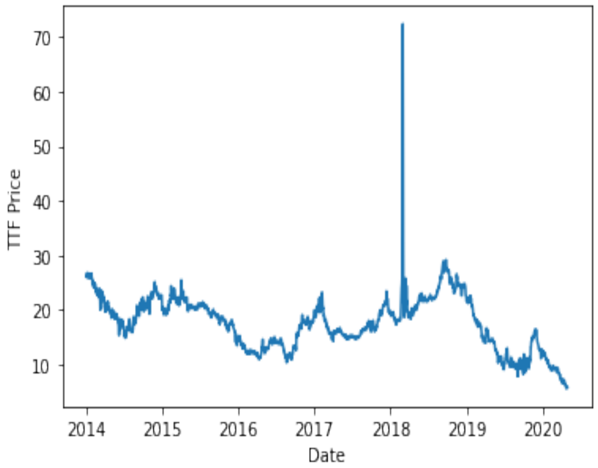

Figure 1 TTF Price Time Series

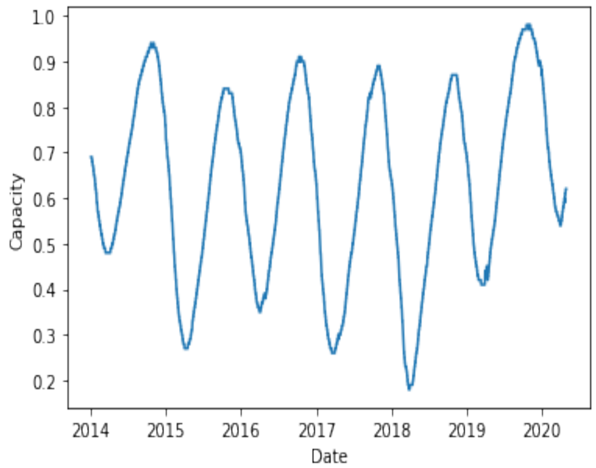

Figure 2 Capacity Time Series

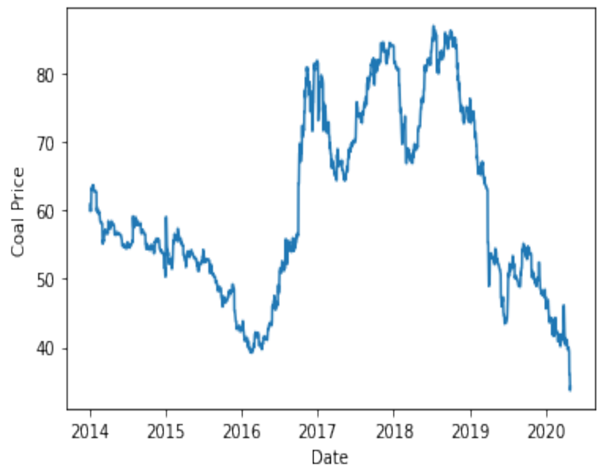

Figure 3 Coal Price Time Series

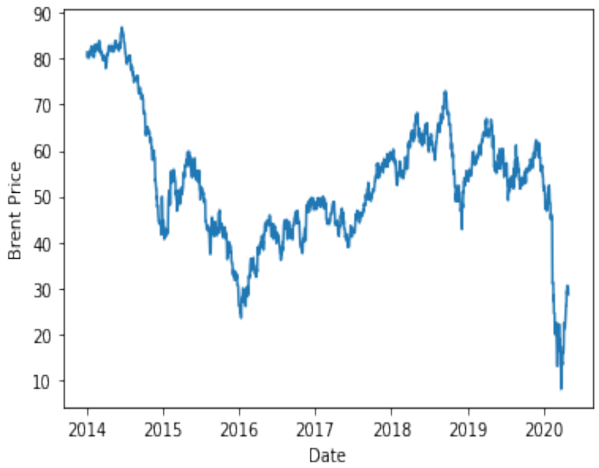

Figure 4 Brent Price Time Series

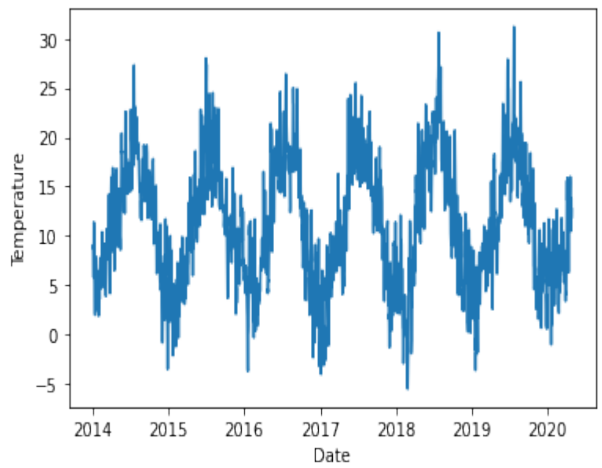

Figure 5 Temperature Time Series

Following the visualization of each variable, it was observed that the "Capacity" (Figure 2) and "Temperature" (Figure 5) time series showed clear seasonal fluctuations, suggesting the presence of regular patterns or cycles. Specifically, the "Temperature" values increased during the Spring and Summer months and decreased during Autumn and Winter. The same pattern was observed in the "Capacity" values. These fluctuations were considered to be natural variations of the time series and were not altered in any way. The sudden spike in the "TTF Price" time series was also prominent and was likely the result of a sudden change in the market. The other plots did not reveal any notable fluctuations or patterns that would require further attention or manipulation.

To further analyse the patterns and trends in the data, descriptive statistics were employed, which involved calculating various measures of central tendency, such as the sum, mean, median, variance, standard deviation, minimum value, and maximum value for each variable. Additionally, to gain insight into the distribution of the data, measures of skewness and kurtosis were also computed. Functions such as "describe", "median", "var", "skew", and "kurt" provided all the aforementioned information.

Last but not least, the time series was checked for correlation and autocorrelation in order to identify potential issues and limitations of the data. While correlation refers to the relationship between two variables, autocorrelation refers to the degree of correlation between a variable and its past values. Through this process, one can identify potential confounding factors and avoid biased estimations. The correlation between variables was examined with the "corr" function and was then visualized with a correlation heatmap. Autocorrelation and Partial Autocorrelation were examined with the "autocorr" and "pacf" functions respectively. These metrics were visualized with the ACF (Auto-Correlation Function) and the PACF (Partial Auto-Correlation Function) plots. With the completion of the EDA process, the next step of the analysis could commence.

5.4.2 Dataset Splitting

Dataset splitting is the process of dividing a dataset into two separate datasets, in order to evaluate the performance of statistical or machine-learning models on new data (Johnson & Kuhn, 2019). It is an important step in model development and data science since it assesses a model's ability to perform well on new observations and avoid overfitting. This process (commonly known as train-test splitting) includes splitting a dataset into two distinct parts, a train set, and a test set. The train set is used to train a specific model and the test set (which contains unseen data) is used to evaluate the model's performance (Johnson & Kuhn, 2019).

Dataset splitting involves the following formula:

$$T = \{(Xi, y1), (X2, y2), \dots, (Xn, yn)\}$$

Where:

- T represents the dataset,
- X is the feature vector, and
- y is the dependent variable.

Then, the dataset is randomly split into two disjoint subsets:

$$T_{train} = \{(X1_{train}, y1_{train}), (X2_{train}, y2_{train}), \dots, (Xn_{train}, yn_{train})\}$$

and

$$T_{test} = \{(X1_{test}, y1_{test}), (X2_{test}, y2_{test}), \dots, (Xn_{test}, yn_{test})\}$$

Where:

- T_train is the training set and
- T_test is the test set.

The common practice for splitting a dataset into training and testing sets is to use a ratio of 70/30 or 80/20. However, the optimal split ratio may differ based on various factors, such as the size and complexity of the dataset, as well as the specific problem and model under consideration.

For this study, the use of machine learning models encourages the use of dataset splitting. Before proceeding with the split, the index of the dataset was reset to avoid errors. The “loc” function was utilised to split the dataset into two, the dependent variable set and the regressor set. Using an 80/20 split, the datasets were split into training sets (containing 1838 observations) and test sets (containing 459 observations). The resulting datasets were the following:

- Dependent_train set: $2\ columns * 1838\ rows$
- Dependent_test set: $2\ columns * 459\ rows$
- Regressors_train set: $5\ columns * 1838\ rows$
- Regressors_test set: $5\ columns * 459\ rows$

After splitting the datasets, the models can be trained with the train set and evaluated with the test set.

5.4.3 Supervised Learning

Supervised learning is a subset of Artificial Intelligence and Machine Learning. Following the “train-test split” process, a model can be trained on a specific dataset, where the inputs and corresponding outputs are known. The primary objective of supervised learning is to develop a mapping function that allows the model to accurately predict outcomes on new data and minimize bias.

Once all four models were fit to the dataset, the next step was to compare their performance using evaluation metrics such as the coefficient of determination (R-squared), root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). These metrics were calculated using the appropriate functions from the "sklearn.metrics" library. The mathematical formula for each evaluation metric is displayed in Table 3.

Table 3 Formulas of Evaluation Metrics

Evaluation Metric	Formula	Evaluation Metric	Formula
RMSE	$\sqrt{\frac{1}{n} \times \sum_{i=1}^{n}(yi - \hat{y}i)^2}$	MAE	$\frac{1}{n} \times \sum_{i=1}^{n}(yi - \hat{y}i)$
MAPE	$\frac{1}{n} \times \sum_{i=1}^{n} \frac{(yi-\hat{y}i)}{yi} \times 100\%$	R-squared	$1 - \frac{\sum(y - \hat{y})^2}{\sum(y - \bar{y})^2}$

SVR

SVR was the first model to be examined. To prepare the data for model training, the “Date” column was set as the index for both the training, and test datasets and the "StandardScaler" function was applied to standardise the regressors. “StandardScaler” is commonly used to preprocess data before model training. Essentially, the function scales the regressors, so that they have a standard deviation of one and a mean of zero. This is done because many machine learning algorithms assume that the regressors are on a similar scale.

Next, the data were split into train and test datasets, and the “GridSearchCV” function was utilised to optimize the hyperparameters of the SVR model. The best hyperparameters were then used to train the SVR model on the training data, and the "predict" function was used to generate predictions of the dependent variable. To evaluate the model's performance, several evaluation metrics such as RMSE, MAPE, MAE, and R-squared were computed.

To improve the accuracy of the model's predictions, first differencing was performed on the data to transform them into a stationary format. Stationarity is a key property of time series data where the statistical properties of the data remain constant over time, such as the mean and variance. This is important because many time series models, including the SVR model, assume stationarity in the data.

After applying the first differencing technique on the data, the optimization and training of the model were repeated. The aim was to evaluate the model’s performance on the stationary data, which has the potential to enhance the accuracy of the model's predictions. The "predict" function was then utilised to generate predictions on the stationary data, and the evaluation metrics were computed to assess the model's performance.

SARIMAX

SARIMAX was the second model to be examined. Before training the model, the dataset’s stationarity was tested with the use of the Augmented Dickey-Fuller (ADF) test. It provided several metrics such as ADF statistics, p-value, number of lags and observations, and critical values. The resulting ADF statistic was -1.9779 with a corresponding p-value of 0.29. Since the p-value was greater than the significance level of 0.05, it was concluded that the time series was non-stationary.

The next step included setting the “Date” column as the index for both dependent and independent variables of the training set and then creating a SARIMAX model with the use of the “auto_arima” function. The model was set to consider seasonal patterns in the data (seasons=True) and to use a stepwise approach (stepwise=True). This would allow the “auto_arima” function to automatically choose the optimal hyperparameters for the model, including the orders of autoregression (AR), integrated (I), and moving average (MA) terms, as well as the seasonal orders for these terms. Additionally, the model was set to allow the function to automatically determine the maximum order of differencing, seasonal differencing, and autoregression (max order=None), and to suppress any warning messages (suppress warnings=True).

The output of the SARIMAX model provided information on the model’s parameters and performance (Table 4). The model selected was SARIMAX(2,1,1) with seasonal components. The AIC and BIC values were 5298.666 and 5320.729 respectively. Additionally, all the coefficients had p-values less than 0.05, indicating that they were statistically significant. To check for the presence of autocorrelation in the residuals, the Ljung-Box test was performed, and the result was not significant, with a p-value of 0.73, indicating that the residuals were not autocorrelated.

Table 4 SARIMAX Results

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                 1838
Model:               SARIMAX(2, 1, 1)   Log Likelihood               -2645.333
Date:                Thu, 30 Mar 2023   AIC                           5298.666
Time:                        19:12:22   BIC                           5320.729
Sample:                             0   HQIC                          5306.802
                               - 1838
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          1.0105      0.008    130.374      0.000       0.995       1.026
ar.L2         -0.3168      0.009    -35.171      0.000      -0.334      -0.299
ma.L1         -0.8814      0.008   -117.465      0.000      -0.896      -0.867
sigma2         1.0428      0.004    268.517      0.000       1.035       1.050
===================================================================================
Ljung-Box (L1) (Q):                   0.12   Jarque-Bera (JB):            9747036.09
Prob(Q):                              0.73   Prob(JB):                          0.00
Heteroskedasticity (H):               8.91   Skew:                             -0.45
Prob(H) (two-sided):                  0.00   Kurtosis:                        359.85
===================================================================================
```

Overall, the output results suggested that the data were a good fit for the SARIMAX model. Once the model was fitted, the “predict” function was utilised to generate forecasts for the dependent variable (Dependent_test) using the independent variables in the test set (Regressors_test). The process concluded by calculating several evaluation metrics of the model, such as RMSE, MAPE, MAE, and R-squared.

MLR

MLR was the third model to be examined. The process was initiated by setting the Regressors and Dependent train sets as “LRX_train” and “LRY_train” respectively. The same process occurred for the test sets (Dependent_test as LRY_test and Regressors_test as LRX_test). In this instance, the model did not require the “Date” to be set as the index, since it was not used as a variable to train or test the model. Then the “LinearRegression” function was used to perform linear regression on the training sets.

After fitting the model, the “predict()” function was utilised to generate forecasts for the dependent variable (LRY_test) using the independent variables (LRX_test). The performance of the model was evaluated by calculating four evaluation metrics, those being RMSE, MAPE, MAE, and R-squared.

RNN+LSTM

The final model to be examined was RNN, using the LSTM technique as the recurrent layer. This choice was made to address the challenges of learning long-term dependencies in sequential data since LSTM is particularly good at overcoming the vanishing gradient problem. The resulting model would be able to selectively remember and forget information over extended periods.

Initially, the “Date” column was set as the index of the data frame. The dependent variable and regressors were set as “target_var” and “regressors” respectively. The data frame was then split into training (1838 observations) and test sets (459 observations), which were scaled with the “MinMaxScaler” function. The “MinMaxScaler” function is a preprocessing technique used in machine learning that transforms the data by subtracting the minimum value of each feature and then dividing it by the range. This was done to ensure that each feature contributed equally to the model's predictions.

Next, the “create_samples” function was defined to transform the time series into a format suitable for supervised learning. Specifically, it created input/output pairs of samples from the time series dataset, making them suitable for the LSTM model. This was done by using time series data and “n_steps” times steps as inputs, and the corresponding value in the series as output. These input-output pairs would be stored in separate lists (X and Y), and then inserted into the LSTM model as “NumPy arrays”. In this instance, the number of time steps was set to five, meaning the model would use the past five observations to predict the following observation.

The LSTM model was created with the use of the “Keras Sequential API” neural network, as an empty object with layers. The first layer included the LSTM model with 100 neurons, the Rectified Linear Unit activation function (ReLu), and the input shape. This was followed by a fully connected “Dense” layer with one neuron, which would produce the predicted values. The output of the LSTM layer was fed into the “Dense” layer, in order to reduce the sequence into a single value. To minimize the difference between the predicted values and the test values, the “Adam” optimization algorithm was used (during training) to update the model’s parameters based on the gradient of the loss function (MSE).

Then the model was fit with the training data for 200 epochs and the “verbose” parameter was set to zero, in order to suppress the output during training. The “predict()” function was utilised to generate forecasts on the test set. The resulting predictions were inverse-transformed using the “MinMaxScaler” function to obtain the actual values. The model's parameters were optimized during training using the "Adam" optimization algorithm, which updates the parameters based on the gradient of the mean squared error (MSE) loss function. The aim was to reduce the difference between the predicted values and the actual test values.

Finally, the model’s true performance on unseen data was estimated by running the model nine more times and then calculating the average of each performance metric.

6. Results

This chapter presents the results of the research study. The results include descriptive statistics of the dataset, analysis of correlation and autocorrelation, and the accuracy of each forecasting model.

6.1 Descriptive Statistics

This section provides an overview of the main characteristics of the data. Specifically, it includes fundamental measures of statistical analysis such as central tendency, variance, and shape.

The first metrics to be analysed, were summary statistics for the dependent variable and regressors. Table 5 contains the main characteristics of the time series.

Table 5 Measures of Central Tendency

	TTF Price	**Temperature**	**Capacity**	**Coal Price**	**Brent Price**
Count	2297.00	2297.00	2297.00	2297.00	2297.00
Mean	17.59	11.20	0.62	60.52	53.22
Min	5.67	-5.50	0.18	33.72	8.00
25%	14.25	6.20	0.47	50.74	43.95
50%	17.82	11.00	0.63	55.95	52.88
75%	21.02	16.10	0.82	72.94	60.31
Max	72.37	31.20	0.98	86.97	86.85

The most significant outcomes of the population analysis were the following:

- The minimum and maximum values of “TTF Price” were 5.67 and 72.37 respectively. The median value was 17.82, indicating that half of the observations were below this value and half were above it. The average value was 17.58.
- The minimum and maximum values of “Temperature” were -5.5 and 31.2 respectively. The average value was 11.2 and the median value was 11.0.
- The minimum and maximum values of “Capacity” were 0.18 and 0.98 respectively. The average value was 0.62. The 75th percentile value for “Capacity” was 0.82, indicating that 75% of the observations were below this value and 25% were above it.

- The minimum and maximum values of "Coal Price" were 33.72 and 86.97 respectively. The median value was 55.95 and the average value was 60.52.
- The minimum and maximum values of "Brent Price" were 8.00 and 86.85 respectively. The median value was 52.88 and the average value was 53.22.

Overall, these measurements provided an overview of the range and distribution of each variable in the time series.

The measures of variance for the time series are included in Table 6.

Table 6 Measures of Variability

	TTF Price	Temperature	Capacity	Coal Price	Brent Price
Std	4.92	6.33	0.20	13.55	14.21
Var	24.24	40.05	0.04	183.70	202.11

Specifically, it included the values of variance and standard deviation for the dependent and independent variables. Based on Table 6, it seems that the values of the "TTF Price" had some variability since they were spread around the mean value. "Temperature" values were more widely dispersed than the "TTF Price" values, indicating more variability in the data. On the other hand, "Capacity" values were relatively tightly clustered around the mean, with little variability present. "Coal Price" and "Brent Price" values had considerable variability, suggesting the presence of extreme values in each column. On the whole, the level of dispersion between values wasn't equal between variables.

The measures of shape for the time series are presented in Table 7.

Table 7 Measures of Shape

	TTF Price	Temperature	Capacity	Coal Price	Brent Price
Skew	7.20	-0.60	-1.04	-1.10	0.22
Kurt	0.69	0.09	-0.18	-0.18	0.16

Particularly, the values of Skewness for each of the dependent and independent variables. "TTF Price" had a high positive skewness value, suggesting that there were more extreme values on the higher end of the range. Conversely, "Temperature" had a slightly negative skewness value, indicating that there were more extreme values at the lower end of the range. "Capacity" and "Coal Price" both had negative skewness values, suggesting that most of the values were on the higher end of the range, with relatively fewer points on the lower end. Finally, "Brent Price" had a small positive skewness value, suggesting more extreme values at the higher end of the range.

Table 7 also included the values of Kurtosis for the dependent variable and regressors. “TTF Price” had a positive kurtosis value, which indicated the presence of outliers. “Temperature” had a slightly positive kurtosis value, meaning that the data were slightly concentrated around the mean. “Capacity” and “Coal Price” both had negative kurtosis values, suggesting that the data were more spread out and less concentrated around the mean. Finally, “Brent Price” had a positive kurtosis value, which suggests a small number of outliers in the data.

6.2 Correlation and Autocorrelation Analysis

The following section discusses the outcomes of the correlation and autocorrelation analyses conducted in this study, which aimed to explore the strength, direction, and temporal patterns of the relationships between the variables examined.

Table 8 illustrates the correlation coefficients among all variable pairs.

Table 8 Correlation Matrix

	TTF Price	**Temperature**	**Capacity**	**Coal Price**	**Brent Price**
TTF Price	1	-0.182388	-0.119519	0.542994	0.476108
Temperature	-0.182388	1	0.102892	-0.013909	0.142245
Capacity	-0.119519	0.102892	1	-0.007307	0.021345
Coal Price	0.542994	-0.013909	-0.007307	1	0.307382
Brent Price	0.476108	0.142245	0.021345	0.307382	1

The correlation coefficients indicated the strength of the relationship between variables. “TTF Price” had a moderate positive correlation with “Coal Price” (0.54) and “Brent Price” (0.48). On the other hand, “TTF Price” had a weak negative correlation with “Temperature” (-0.18) and “Capacity” (-0.12). Additionally, “Temperature” and “Capacity” had a weak positive correlation (0.10). Finally, “Coal Price” had a moderate positive correlation with “Brent Price” (0.31).

The aforementioned results were also evident through a correlation heatmap, which is a visual representation of the correlation matrix (Figure 6). In this type of graph, the correlation coefficients between all pairs of variables are shown as coloured cells. Lighter colours indicated a lower correlation between variables, while darker colours indicated the opposite.

Darker-coloured cells can be seen between the following pairs: “Coal Price” – “TTF Price”, “Brent Price” – “TTF Price”, and “Brent Price” – “Coal Price”. Light-coloured cells can be seen between “Capacity” and “Temperature”.

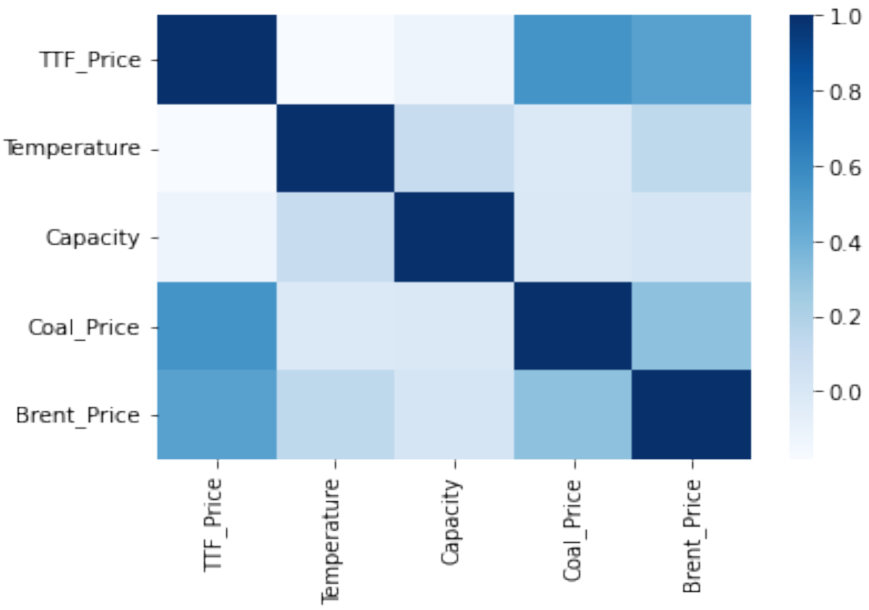

Figure 6 Correlation Heatmap

Autocorrelation was examined with the use of ACF and PACF plots at 20 lags. A lag refers to the number of time periods between two data points that are being compared for correlation. The ACF plot (Figure 7) displays the correlation coefficients between the time series and its lagged values. Like the ACF plot, the PACF plot (Figure 8) presents the correlation between a time series and its lagged values, after accounting for the effects of the intervening observations.

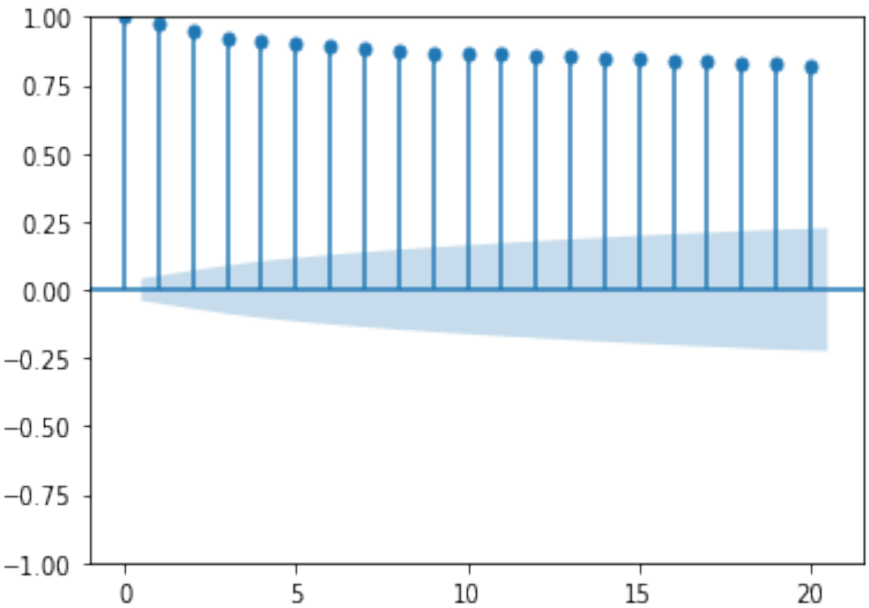

Figure 7 Autocorrelation Plot (ACF)

Figure 7 represents the autocorrelation values for the dependent variable ("TTF Price"). The number of lags was set to 20. Each vertical line matches the correlation coefficient between the original time series and a lagged version of itself at the corresponding lag. The values gradually decrease, indicating a decreasing correlation between the dependent variable and its past values as the lag increases. The fact that the values were still relatively high even after 20 lags, suggested that there was strong autocorrelation in the data. This result may indicate the presence of a trend or seasonality.

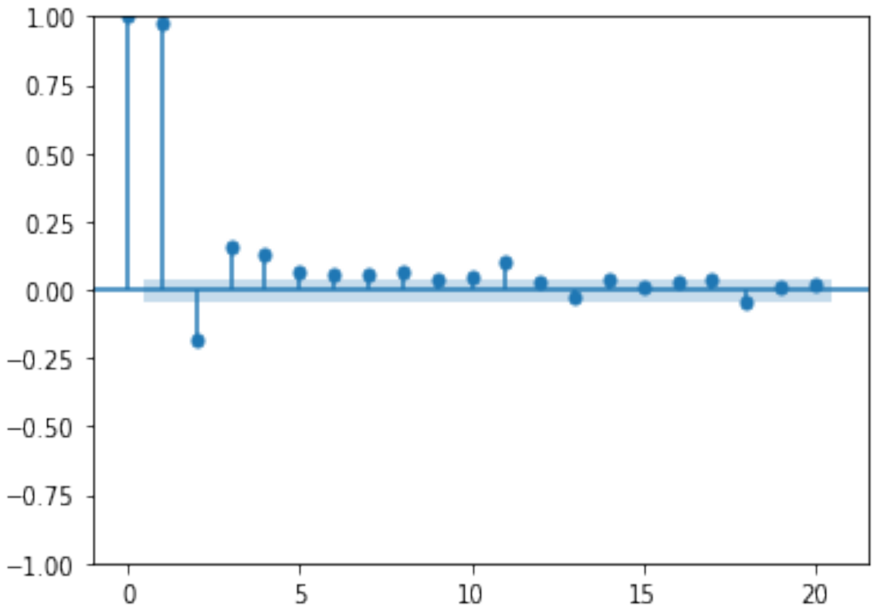

Figure 8 Partial Autocorrelation Plot (PACF)

Figure 8 represents the sequence of autocorrelation coefficients for the dependent variable ("TTF Price"). Each dot matches the partial autocorrelation coefficient of the time series at a specific lag. As the lag increases, partial autocorrelation decreases, with some occasional spikes along the way. The dependent variable was influenced by its past values up to a certain point, and then the correlation became insignificant or close to zero. These spikes may indicate the presence of seasonal patterns in the data.

Overall, both autocorrelation and partial autocorrelation plots suggest that there was a strong linear relationship between adjacent observations in the dependent variable.

6.3 Model Evaluation

The final section presents the evaluation metrics of each forecasting model. Firstly, it is important to note that while the evaluation metrics of MLR and SARIMAX were consistent during each run of the algorithm, RNN+LSTM and SVR evaluation metrics varied during each run. This phenomenon is attributed to the fact that machine learning models such as SVR and RNN+LSTM are stochastic in nature and use random sampling techniques during the training. To obtain a more accurate evaluation of each model's performance, they were run ten times, and the mean value of each metric was calculated.

The evaluation metrics of each model are summarised in Table 9.

Table 9 Evaluation Metrics

	SVR	**SARIMAX**	**MLR**	**RNN+LSTM**
RMSE	0,2958	10.256	5.992	7,5103
MAPE	1,1949	0.955	0.538	0,6625
MAE	0,1607	9.727	5.728	7,099
R-squared	0,9087	-9.004	-2.414	-4,7858

First of all, SVR had the lowest MAE and RMSE values, a relatively high MAPE value, and a high value. On the other hand, SARIMAX had a low MAPE value, very high RMSE and MAE values, and a negative R-squared value. Similarly, MLR had a low MAPE value, relatively high RMSE and MAE values, and a negative R-squared value. Finally, RNN+LSTM had a lower MAPE value than MLR but higher RMSE and MAE values. It also had a negative R-squared value.

Each model had its own unique strengths and weaknesses. The choice of the most suitable model would depend on the specific context and the objectives of the analysis.

7. Conclusions

This chapter illustrates the conclusions drawn from the results of the analysis and forecasting processes.

Based on the results of the analysis, SVR had the lowest MAE and RMSE values, indicating that it had the best performance in predicting values compared to the other models. However, it had a relatively high MAPE, which suggested less accuracy when larger errors occurred. SARIMAX had a low MAPE value, high RMSE and MAE values, and a negative R-squared value, indicating that it had poor performance in predicting values compared to the other models. MLR had a low MAPE value, relatively high RMSE and MAE values, and a negative R-squared value, indicating that it had poorer performance in predicting values compared to SVR. RNN+LSTM had a lower MAPE than MLR, higher RMSE and MAE values, and a negative R-squared value, indicating that it had poorer performance in predicting values compared to SVR and MLR, but better performance compared to SARIMAX.

When considering the forecasting capabilities of the models, the results indicated that the SVR model performed the best, followed by MLR, RNN+LSTM, and ultimately SARIMAX. More specifically, SVR performed the best in terms of predicting the actual values of the dependent variable, but its percentage error was higher than the other models. MLR and RNN+LSTM had similar performance, while SARIMAX performed the worst among the four models.

8. Discussion

The aim of this study was to conduct a comparative analysis of four widely used forecasting models, in order to identify the most accurate model for forecasting natural gas prices in Western Europe. The models selected for the study were MLR (Multiple Linear Regression) and SARIMAX (Seasonal Auto-Regressive Integrated Moving Average with eXogenous factors), which are traditional statistical and time series models, and RNN (Recurrent Neural Network) with Long-Short Term Memory(LSTM) and SVR (Support Vector Regression), which are advanced machine learning models.

The models were trained on historical data and tested on more recent data, and the variable of interest was the TTF Spot Last Price. The study also included several regressors, those being the average temperature of Western Europe, the available natural gas in European storage, the Brent Crude Oil Spot last price (FOB), and the Coal (API2) CIF ARA Continuous Contract last price.

The comparison of evaluation metrics suggested that SVR outperformed traditional models such as MLR and SARIMAX, both in terms of RMSE and MAE metrics. This finding is contrasted by the work of Čeperić et al. (2017), who found that AR and ARIMA outperformed SVR in forecasting natural gas prices. This fact could possibly be attributed to the difference between data and markets since Čeperić et al. (2017) examined the Henry Hub spot price and the U.S. commodities market.

Additionally, RNN+LSTM outperformed SARIMAX in terms of all evaluation metrics but scored lower than MLR. This finding is contrasted by the studies of Azadeh et al. (2012) and Salehnia et al. (2013), who found that Neural Network models outperformed Regression models when forecasting natural gas prices. This observation could also be attributed to the difference between data and markets, since Azadeh et al. (2012) and Salehnia et al. (2013) chose economic indicators from the Iranian and U.S. markets respectively, as the regressors for their research. Another noteworthy finding was that RNN+LSTM scored lower than SVR in terms of RMSE, MAE, and R-squared, which is also contrasted by the findings of Su et al. (2019). Their research concluded that Neural Networks outperformed Support Vector Machines when forecasting natural gas prices. This result could be linked to the higher number of variables and the smaller number of observations utilised in their study.

In regards to the first hypothesis, complex machine learning models did not necessarily outperform traditional statistical and time series models. While advanced machine learning models such as RNN+LSTM and SVR have been shown to perform well in certain contexts, the present study found that traditional statistical models such as MLR can also provide accurate forecasts of natural gas prices. The MLR model may have captured key underlying patterns and relationships in the data, thus being better than the RNN+LSTM model in forecasting natural gas prices.

In reference to the second hypothesis, RNN+LSTM did not outperform the rest of the models in terms of forecasting natural gas prices. The RNN model can be sensitive to the size and quality of the dataset used for training. The dataset used in the analysis may not have been large enough or varied enough to fully capture the complexity of the underlying patterns, leading to suboptimal performance of the RNN+LSTM model.

Overall, natural gas price forecasting can be a complex task, and the choice of the most appropriate forecasting model may depend on the specific characteristics of the data being analysed. Additional research is required to evaluate the accuracy of alternative forecasting models and the significance of natural gas price determinants.

9. Limitations

This chapter acknowledges the limitations of the study, which are important to understand the potential implications of the findings.

The objective of this study was to conduct a comparative analysis of four distinct forecasting models by utilising a number of regressors to forecast natural gas prices in Western Europe. As with any research, there were limitations in building the ideal forecasting model, choosing the best available data, and interpreting the results with the highest possible accuracy.

The first limitation revolved around the time frame of the available data. The data provided by the EEX were limited and did not cover a substantial time period. What's more, the period from 2014 to early 2020 did not include any significant global events that could have affected the price of natural gas, such as geopolitical crises, environmental disasters, or extreme economic regulations and shifts. Therefore, it was not possible to analyse the long-term trends and the impact of external shocks on the West European natural gas market.

Another limitation was the insufficient data on the drivers of natural gas prices. Data on electricity consumption, heating oil and carbon credit prices, and imports of liquid natural gas, were not available at a daily frequency or in a European context. The limited number of regressors could have impeded the abilities of the models and potentially biased the results.

Missing values were another possible limitation of the analysis. The available data had missing values which were addressed through the use of mean imputation. While this is a common method for dealing with missing data, it may have introduced some errors or biases in the analysis.

An additional limitation of this study was the limited number of models used to analyse the data, as the choice of models was restricted by time limitations. This may have resulted in an incomplete examination of the optimal forecasting model and could have potentially overlooked other models that may have better performance.

The final limitation pertained to the fact that the models were not tested with each variable individually. This approach could have potentially provided valuable insights into the relative importance of each variable in influencing TTF Spot natural gas prices. This, in turn, would have improved the model's capabilities and increased the accuracy of price forecasts.

To enhance the precision and dependability of the analysis, future studies could concentrate on overcoming these limitations.

10. Suggestions for Future Research

This chapter will present research proposals in order to improve the precision and reliability of forecasting models. There are several aspects that can be improved with future research.

Conducting research with more variables such as geopolitical events, economic indicators, and technological advancements, could improve the accuracy and reliability of forecasting models. A longer time frame of available data could also provide a better understanding of natural gas price dynamics.

Similarly, exploring and evaluating more modelling techniques such as advanced machine learning and ensemble methods could provide insights into non-linear relationships and interactions between variables.

Additionally, future research could also focus on testing the models with each variable individually to gain more insight into their respective impacts on natural gas prices.

Furthermore, addressing missing values and price spikes in a more robust manner, such as using advanced imputation techniques or outlier detection methods, would help reduce potential bias and increase the robustness of the analysis.

Finally, research should be extended to other natural gas markets. A researcher could utilise similar variables and models, to conduct analyses of the Asian and American markets. This in turn could lead to a global approach when forecasting natural gas prices.

References

Abe, S. (2005). *Support vector machines for pattern classification* (2nd ed.). Springer.

Ali, A. (2020). Ensemble learning model for prediction of natural gas spot price based on least squares boosting algorithm. *In 2020 International conference on data analytics for business and industry: way towards a sustainable economy (ICDABI)*. Sakhir, Bahrain 26-27 Oct. 2020. IEEE Xplore.

Aminu, N. (2019). Energy prices volatility and the United Kingdom: Evidence from a dynamic stochastic general equilibrium model. *Energy*, *172*, 487-497.

Atwater, G. I., Solomon,. L. H., Carruthers, J. E., Riva, J. P., & Waddams A. L. (2023). natural gas. In *Encyclopedia Britannica*. https://www.britannica.com/science/natural-gas.

Avouris, N., Koukias, M., Paliouras, V., & Sgarbas, K. (2020). *Python: Introduction to Computers* (4th ed.). University Publications of Crete.

Azadeh, A., Sheikhalishahi, M., & Shahmiri, S. (2012). A hybrid neuro-fuzzy approach for improvement of natural gas price forecasting in vague and noisy environments: domestic and industrial sectors. In *Proceedings of the International Conference on Trends in Industrial and Mechanical Engineering* (ICTIME'2012). Dubai, United Arab Emirates 24-25. 2012.

Batten, J. A., Ciner, C., & Lucey, B. M. (2017). The dynamic linkages between crude oil and natural gas markets. *Energy Economics*, *62*, 155–170.

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, *5*(2), 157-166.

Bianchi, F. M., Maiorino, E., Kampffmeyer, M. C., Rizzi, A., & Jenssen, R. (2017). *Recurrent neural networks for short-term load forecasting: an overview and comparative analysis*. Springer.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Broadstock, D. C., Li, R., & Wang, L. (2020). Integration reforms in the European natural gas market: A rolling-window spillover analysis. *Energy Economics*, *92*, 104939.

Brockwell, P. J., & Davis, R. A. (2002). *Introduction to time series and forecasting* (2nd ed.). Springer.

Brown, S. P. A., & Yucel, M. K. (2008). What drives natural gas prices?. *The Energy Journal*, *29*, 45–60.

Čeperić, E., Žiković, S., & Čeperić, V. (2017). Short-term forecasting of natural gas prices using machine learning and feature selection algorithms. *Energy*, *140*, 893-900.

Chen, J. (2022). Brent Blend. In J. Mansa (Ed.), *Investopedia*. https://www.investopedia.com/terms/b/brentblend.asp.

Chen, Y., Hartley, P. R., & Lan, Y. (2023). Temperature, storage, and natural gas futures prices. *Journal of Futures Markets*, *43*(4), 549-575.

Chestney N. (2022). Explainer: How natural gas is traded in Europe. *Reuters.* https://www.reuters.com/business/energy/how-natural-gas-is-traded-europe-2022-12-20.

Corbett, R. A. (1983). Guide to world export crudes Brent blend, North Sea benchmark, assayed. *Oil and Gas Journal*, *88*(26), 69-76.

Erdogdu, E. (2010). Natural gas demand in Turkey. *Applied Energy*, *87*(1), 211-219.

Fernández Alvarez, C. (2022). The Trading and Price Discovery for Coal. In M. Hafner & G. Luciani (Eds.), *The Palgrave Handbook of International Energy Economics* (pp. 395-406). Springer Nature AG.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). Sage.

Gay, G. D., Simkins, B. J., & Turac, M. (2009). Analyst forecasts and price discovery in futures markets: The case of natural gas storage. *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, *29*(5), 451-477.

Geron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems* (2nd ed.). O'Reilly Media, Inc..

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.

Gujarati, D. N., & Porter, D. C. (2009). *Basic econometrics* (5th ed.). McGraw-Hill.

Hafner, M. & Luciani, G. (2022). The Trading and Price Discovery for Natural Gas. In M. Hafner & G. Luciani (Eds.), *The Palgrave Handbook of International Energy Economics* (pp. 377-394). Springer Nature AG.

Hafner, M., & Raimondi, P. P. (2022). Energy and the Economy in Europe. In M. Hafner & G. Luciani (Eds.), *The Palgrave Handbook of International Energy Economics* (pp. 731-761). Springer Nature AG.

Halova, M. W., Kurov, A., & Kucher, O. (2014). Noisy inventory announcements and energy prices. *Journal of Futures Markets, 34*(10), 911-933.

Hartley, P. R., & Medlock, K. B., III (2014). The relationship between crude oil and natural gas prices: The role of the exchange rate. *The Energy Journal*, *35*(2), 25–44.

Hartley, P. R., Medlock, K. B., III, & Rosthal, J. E. (2008). The relationship of natural gas to oil prices. *The Energy Journal, 29*(3), 47–65.

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). T*he elements of statistical learning: data mining, inference, and prediction* (2nd ed.). Springer.

Honoré, A. (2013). *The Italian Gas Market–Challenges and Opportunities*. Oxford Institute for Energy Studies.

Hosseinipoor, S., Hajirezaie, S., & Nejati, J. (2016). Application of ARIMA and GARCH Models in Forecasting the Natural Gas Prices. The *University of Oklahoma*.

Hulshof, D., Van Der Maat, J. P., & Mulder, M. (2016). Market fundamentals, competition and natural-gas prices. *Energy policy, 94*, 480-491.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice* (3rd ed.). OTexts.

Jeon, J. (2015). The strengths and limitations of the statistical modeling of complex social phenomenon: Focusing on SEM, path analysis, or multiple regression models. *International Journal of Economics and Management Engineering*, *9*(5), 1634-1642.

Jin, J., & Kim, J. (2015). Forecasting natural gas prices using wavelets, time series, and artificial neural networks. *PloS one*, *10*(11), e0142064.

Johnson, K., & Kuhn, M., (2019). *Feature Engineering and Selection: A Practical Approach for Predictive Models.* CRC Press.

Jotanovic, V., & D'Ecclesia, R. L. (2021). The European gas market: new evidences. *Annals of Operations Research*, *299*(1-2), 963-999.

Kagan, J. (2021). Continuous Contract. In D. Kindness (Ed.), *Investopedia*. https://www.investopedia.com/terms/c/continuous-contract.asp.

Kalogiannidis, S., Chatzitheodoridis, F., Kalfas, D., Kontsas, S., & Toska, E. (2022). The Economic Impact of Russia's Ukraine Conflict on the EU Fuel Markets. *International Journal of Energy Economics and Policy*, *12*(6), 37-49.

Kuper, G. H., & Mulder, M. (2016). Cross-border constraints, institutional changes and integration of the Dutch–German gas market. *Energy Economics*, *53*, 182-192.

Kurt, D. (2022). Benchmark Oils: Brent Crude, WTI and Dubai. In C. Potters (Ed.), *Investopedia*. https://www.investopedia.com/articles/investing/102314/understanding-benchmark-oils-brent-blend-wti-and-dubai.asp.

Li, H., Chen, L., Wang, D., & Zhang, H. (2017). Analysis of the price correlation between the international natural gas and coal. *Energy Procedia*, *142*, 3141-3146.

Li, R., & Song, X. (2023). A multi-scale model with feature recognition for the use of energy futures price forecasting. *Expert Systems with Applications*, *211*, 118622.

Lin, B., & Li, J. (2015). The spillover effects across natural gas and oil markets: based on the VEC–MGARCH framework. *Applied Energy*, *155*(1), 229–241.

Linn, S. C., & Zhu, Z. (2004). Natural gas prices and the gas storage report: Public news and volatility in energy futures markets. *Journal of Futures Markets*, *24*(3), 283–313.

Liu, B., & Geman, H. (2017). World coal markets: Still weakly integrated and moving east. *Journal of commodity markets*, 5, 63-76.

Lv, X., & Shan, X. (2013). Modeling natural gas market volatility using GARCH with different distributions. *Physica A: Statistical Mechanics and its Applications*, *392*(22), 5685-5699.

Mosquera-López, S., & Nursimulu, A. (2019). Drivers of electricity price dynamics: Comparative analysis of spot and futures markets. *Energy Policy*, *126*, 76-87.

Mu, X. (2007). Weather, storage, and natural gas price dynamics: Fundamentals and volatility. Energy Economics, 29(1), 46-63.

Muller, A., & Guido, S. (2016). *Introduction to machine learning with Python: A guide for data scientists*. O'Reilly Media, Inc..

Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. MIT Press.

Nick, S., & Thoenes, S. (2014). What drives natural gas prices?—A structural VAR approach. *Energy Economics*, *45*, 517-527.

Papież, M., Rubaszek, M., Szafranek, K., & Śmiech, S. (2022). Are European natural gas markets connected? A time-varying spillovers analysis. *Resources Policy*, *79*, 103029.

Richter, P. M., & Holz, F. (2015). All quiet on the eastern front? Disruption scenarios of Russian natural gas supply to Europe. *Energy Policy*, *80*, 177-189.

Roinioti, A., Simitchiev, R., Sofianos, N., Yardim, G., Lee, J., Manis, P., Tsotsos, R., (2014). The Outlook for a Natural Gas Trading Hub in Europe. In C. Stambolis (Ed.), *Institute of Energy for South-East Europe.*

Rubaszek, M., & Uddin, G. S. (2020). The role of underground storage in the dynamics of the US natural gas market: A threshold model analysis. *Energy Economics*, *87*, 104713.

Salehnia, N., Falahi, M. A., Seifi, A., & Adeli, M. H. M. (2013). Forecasting natural gas spot prices with nonlinear modeling using Gamma test analysis. *Journal of Natural Gas Science and Engineering*, *14*, 238-249.

Schmidt, A. F., & Finan, C. (2018). Linear regression and the normality assumption. *Journal of clinical epidemiology, 98*, 146-151.

Seo, S. H. (2021). Forecasting Korean LNG import price using ARIMAX, VECM, LSTM and hybrid models. *Ulsan National Institute of Science & Technology*.

Shi, X., & Variam, H. M. (2018). Key elements for functioning gas hubs: A case study of East Asia. *Natural Gas Industry B*, *5*(2), 167-176.

Shumway, R. H., & Stoffer, D. S. (2011). *Time series analysis and its applications: with R examples* (3rd ed.). Springer.

Su, M., Zhang, Z., Zhu, Y., Zha, D., & Wen, W. (2019). Data driven natural gas spot price prediction models using machine learning methods. *Energies*, 12(9), 1680.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson.

Tong, X., Zheng, J., & Fang, B. (2014). Strategic analysis on establishing a natural gas trading hub in China. *Natural Gas Industry B*, *1*(2), 210-220.

U.S. Department of Energy, Energy Information Administration, Independent Statistics & Analysis. (2016). *Glossary*. https://www.eia.gov/tools/glossary/.

U.S. Department of Energy, Energy Information Administration, Independent Statistics & Analysis. (2022). *Natural gas explained.* https://www.eia.gov/energyexplained/natural-gas/.

Wang, S., & Chaovalitwongse, W. A. (2011). Evaluating and comparing forecasting models. *Wiley Online Library.*

Wang, T., Zhang, D., & Broadstock, D. C. (2019). Financialization, fundamentals, and the time-varying determinants of US natural gas prices. *Energy Economics*, *80*, 707-719.