Johan Sæbø

An Analysis of UK Natural Gas Prices Comparing Regression, Neural Nets and Gradient Boosting Techniques

Master's thesis in Applied Physics and Mathematics
Supervisor: Jacob Kooter Laading
June 2023

Master's thesis

NTNU
Norwegian University of
Science and Technology

Johan Sæbø

An Analysis of UK Natural Gas Prices Comparing Regression, Neural Nets and Gradient Boosting Techniques

Master's thesis in Applied Physics and Mathematics
Supervisor: Jacob Kooter Laading
June 2023

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

NTNU
Norwegian University of
Science and Technology

Abstract

This master thesis presents a comparative analysis of three different modeling methods: multiple linear regression, the gradient boosting technique XGBoost, and the neural network Long Short-Term Memory (LSTM). To compare the models, the financial time series of the natural gas price in the UK was used.

The main goal of this investigation was to explore the effectiveness of different modeling approaches in capturing the complex relationships between variables that impact the natural gas price and assess their predictive performance on a highly volatile commodity such as the gas price. These models represent distinct branches of modeling, including traditional statistical modeling, machine learning, and neural networks.

Two different periods were selected to assess the models' efficacy, one period during a stable market between 2018 and the beginning of 2020, and the other during a market crisis that started with the 2021 natural gas supplier crisis in the UK and continued with the Russian invasion of Ukraine. The data covered the period from the beginning of 2009 to the end of 2022.

During the "normal" period, all three models performed adequately in terms of prediction, with the multiple linear regression model showing a slightly weaker performance than the other two. The LSTM model marginally outperformed XGBoost. During the highly volatile period, the multiple linear regression model demonstrated success in capturing the price dynamics and did successfully identify two significant price surges, out of the three observed.

In terms of overall prediction accuracy, the LSTM model exhibited the best performance, with the XGBoost model a close second, and the multiple linear regression model exhibited a comparatively lower performance. However, both LSTM and XGBoost struggled to extrapolate the price surges but still demonstrated a strong overall fit.

Sammendrag

I denne masteroppgaven ble det utført en komparativ analyse av de tre ulike modeller-
ingsmetodene: multippel lineær regresjon, gradient boosting metoden XGBoost og det
nevrale nettverket Long Short-Term Memory (LSTM). For å sammenligne modellene ble
tidsserie data for naturgassprisen i UK brukt.

Målet med denne masteroppgaven var å utforske i hvilken grad de ulike modelleringsmet-
odene klarte å konstruere sammenhenger mellom variabler, som påvirker naturgassprisen,
og vurdere deres prediksjonsytelse for en svært volatil handelsvare som gassprisen. Disse
tre modellene representerer forskjellige grener av modellering som tradisjonell statistisk
modellering, maskinlæring og nevrale nettverk.

To ulike perioder ble valgt for å vurdere modellene: en periode under en stabil markeds-
situasjon mellom 2018 og begynnelsen av 2020, og en annen periode under en markedskrise
som startet med gassleverandørkrisen i Storbritannia i 2021 og fortsatte med Russlands
invasjon av Ukraina. Dataene dekker perioden fra begynnelsen av 2009 til slutten av
2022.

I den "normale" perioden presterte alle tre modellene tilfredsstillende med hensyn til
prediksjon, der multiple lineær regresjonsmodell viste en noe svakere ytelse enn de to
andre. LSTM-modellen presterte marginalt bedre enn XGBoost. I den svært volatile
perioden lyktes regresjonsmodellen med å fange opp pris-dynamikken og identifiserte to
av de tre store pris-hoppene.

For den gjennomsnittlige prediksjonsnøyaktigheten var LSTM-modellen best, med XGBoost-
modellen like bak og sist, regresjonsmodellen. Både LSTM og XGBoost hadde problemer
med å ekstrapolere de bratte prisstigningene, men viste fortsatt en god tilpasningsevne.

Preface

It is with great pride and satisfaction that I present this master thesis, which serves as the culmination of my master degree in Physics and Mathematics with a specialisation in Industrial Mathematics. The master thesis is written in the course TMA4900 during the spring of 2023 at the Norwegian Institute of Science and technology(NTNU). This research endeavor has been a remarkable and transformative experience, allowing me to delve deep into the intricacies of my chosen field.

I would like to express my sincere gratitude to my supervisor, Jacob Kooter Laading, for his guidance, support, and insights throughout writing this master thesis starting in the fall of the Autumn of 2022. His experience and knowledge have been important in shaping this thesis and my overall academic growth.

Lastly, I wish to extend my sincere appreciation to my friends, family and girlfriend whose encouragement, and belief in my abilities have been the cornerstone of my academic pursuits. Their constant support and help have provided me with the necessary motivation to overcome challenges and strive for excellence.

Table of Contents

List of Figures

List of Tables

1 Introduction

This paper aims to compare a Machine Learning (ML) model, a Neural Network, and a multiple linear regression to examine how they perform in a highly volatile market such as the natural gas market, affected by numerous factors.

Energy has been vital for human civilization for millennia. After the industrial revolution, the world has seen an exponential growth in its population thanks to new technological advances, driven by energy. Since then, the value and utility of energy has increased accordingly with the development of innovative technology.

In recent years, there has been an increase of awareness of the problems of some energy sources regarding climate change and the vulnerability of being too dependent on other countries. Nevertheless, as of today the world is still heavily reliant on fossil energy. One of these vital fossil energy sources is natural gas. The natural gas price in a developed market changes according to the normal financial principle "supply and demand", but it has not always been the case[6].

Before 1996 all European natural gas markets were indexed against the brent crude oil price. Since then, these markets have transitioned towards free and open markets. Consequentially, the natural gas market is evolving every year with external factors, making the natural gas market difficult to predict. Nevertheless, due to the importance of natural gas, there are great benefits to gain from a better understanding of the dynamics of the price movements.

This paper is structured as follows: Chapter 2 provides a comprehensive examination of the natural gas market, with its historical development and the underlying factors influencing price dynamics. Chapter 3 presents the data set used in this study and offers descriptive statistics of the covariates. In Chapter 4, the field of financial time series modeling is introduced, equipping readers with the necessary theoretical foundations for subsequent analyses. Chapters 5, 6, and 7 delve into the three individual models, explaining their respective theories, algorithms, and implementation procedures. Chapter 8 presents the results of the models, with a discussion of the models performance, similarities, and differences. The last chapter concludes the study's findings and outlines future research directions. This thesis builds to some extent on my work in the TMA4500 project paper, especially chapters 5 and 6.

2 The gas market

To model a market, it is essential to have knowledge of the market. Regarding the UK natural gas market, this should include where and how the gas is imported, what the gas is being used for and other useful information.

2.1 The beginning of the European gas market

Prior to 1996, the European gas market was nonexistent and comprised of national components, which were isolated from each other[6]. The dispensation of gas primarily relied on long-term contracts, with prices indexed against the oil price. To establish a competitive and efficient market, the European Union (EU) started the process of liberalizing the European gas market, resulting in the emergence of different gas hubs and pricing models based on the principles of supply and demand across Europe.

The first European gas hub called the National Balancing Point (NBP) was established in the UK in 1996. The NBP allows the uniform trading of all gas in the UK at a single price via ICE Futures Europe. Notably, the NBP is a virtual hub that lacks physical infrastructure for the physical distribution of gas via pipelines and valves.

In contrast, the Henry Hub located in Erath, Louisiana, is a physical hub that serves as a significant benchmark for the gas market in North America. Since 1990, it has been pivotal in contract deliveries for trading on the NYMEX. The Henry Hub's influence extends beyond North America as the United States has significantly increased its liquefied natural gas (LNG) exports in recent years, thus impacting the global LNG market.

2.2 Logistics of gas transportation

There are 13 Local Distribution Zone (LDZ) which functions as gas distribution networks in the UK, however they are equally priced.
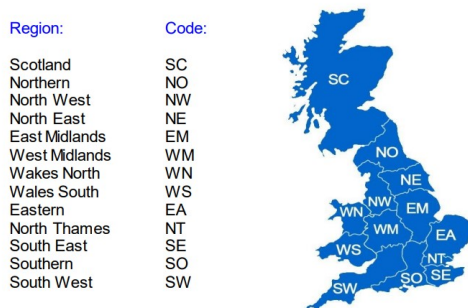

Figure 1: The LDZ in the UK

Source: [41]

The transportation of natural gas is a fundamental requirement in trading between coun-

tries, as well as within individual LDZs in the UK. The construction of a network of gas pipelines throughout Europe, including the UK, facilitates this transportation from buyer to seller. In theory, these pipelines have the capacity to transport unlimited quantities of gas by increasing the velocity of the gas[21]. However, this approach is unviable due to erosion caused by the presence of dust particles in the pipeline. As the velocity of the gas increases, so does the erosion rate, leading to limitations on gas transportation within a given market.

A relative new approach to gas transportation involves transportation by ships between geographically separated markets [42]. This is achieved by cooling the gas to -162 degrees Celsius in order to get LNG. The advantage of liquefaction lies in the significant reduction in volume, thereby increasing the amount of natural gas that can be transported, consequentially increasing the profitability. Increasing the capability of LNG-shipments will reduce the difference in the gas price between markets, but there are limiting factors for this including the availability of ships and the duration of transportation. The LNG sector has become a pivotal component of the global natural gas market.

2.3 Key information about the gas market

2.3.1 Measuring unit of natural gas

The transportation of LNG illustrates a problem when trying to quantify amounts of gas. Since gas can be compressed and expanded it does not make sense to measure quanta in volume-units in the same way as for oil. Instead, the norm is the Metric Million British thermal unit (MMBtu) and is defined as how much energy that is required to raise the temperature of 1 pound of water by 1 degree Fahrenheit at 1 atmospheric pressure-level [39]. However, most natural gas companies operates with the unit therm (Thm) which is converted as 1 MMBtu = 10 Thm [9].

2.3.2 The two main options for traders

For the different hubs mentioned there are mainly two options to trade: spot price or futures contracts[6]. The spot price is a purchase and consumption of gas at the current moment in time. A futures contract is an agreement between a buyer and seller of a given price, but with a later delivery time. This would be beneficial if one wants certainty, but it is also heavily linked to market speculations.

2.3.3 Differences between geographically separated markets

Natural gas markets are geographically restricted with their pipelines, and therefore major price-differences in gas prices can arise between geographically separated markets[13]. With the recent growth in the LNG sector, there is now an increasing possibility to transport gas between markets. However, it is not easy to take advantage of these arbitrage opportunities due to multiple factors. Technical, contractual and market restrictions, together with liquefying, re-gasification, transport-ship capacity and high transportation costs are some of the barriers deterring arbitrage. As there is money to be made, the

LNG sector has grown rapidly and is expected to continue to develop by minimizing costs and expanding its transportation capacity, leading to more gas price convergence between world markets.

2.4   Imports of gas to the UK

The two main contributors to the UK gas supply mix are domestic production with 48% and imports from Norway with 29%, as seen in Figure 2 [3]. Even though the UK has a solid gas reserve in the North Sea to cover its gas demand, this would according to some estimates make the reserve deplete by 2030. The UK government chooses to import gas from other nations, since they see the value of having gas in the ground as a storage [3].

The remaining supply to the UK comes from LNG import from Qatar, USA, Russia, and others. These numbers are from 2020 which explains why Russia was a part of the gas supply mix. The gas price in the UK continued to rise after Russia cut out its export to Europe, despite the low percentage imported. This is due to natural gas being a global commodity, where the prices are given by supply and demand. Since Russia provided a large amount of the world's gas production, the supply to the world gas market took a heavy fall after Russia stopped exporting to Europe. Another example is the USA, which did not import from Russia or Europe but still had a significant increase in gas prices [20].
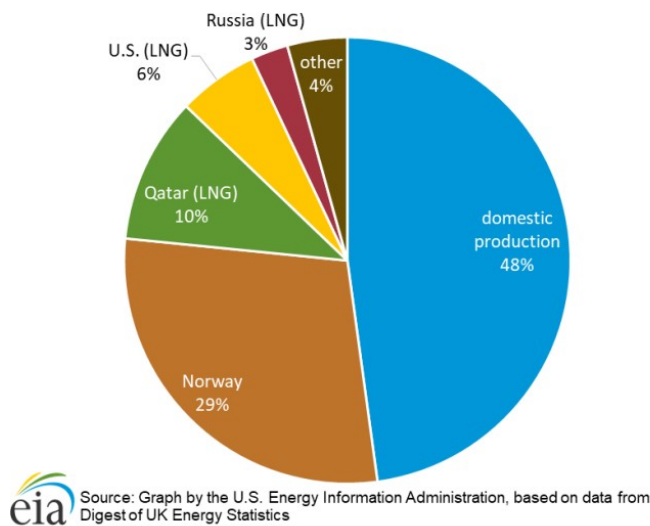

Source: Graph by the U.S. Energy Information Administration, based on data from Digest of UK Energy Statistics

Figure 2: The natural gas import to the UK in GBP in 2020.

Source: [3]

2.5 Factors affecting the gas price

2.5.1 Politics

The importance of reliable and affordable energy cannot be overstated, as it is vital for both households and industries. Consequently, energy markets, including the gas market, are significantly influenced by governmental policies, both directly and indirectly. In recent years, the need to address climate change has led to a increased priority on green energy sources, resulting in more government investments and initiatives across Europe. This shift aims to decrease reliance on carbon-emitting energy, including natural gas, with the ultimate goal of being carbon neutral. Moreover, political decisions, such as the European Union's recent move to cease gas imports from Russia, further exemplify how governmental actions impact the dynamics and pricing of natural gas.

2.5.2 Gas storage and seasonality

Natural gas is used for heating in a large part of Europe, resulting in a significant correlation between gas demand and temperature fluctuations[36]. As a consequence, numerous countries fill their gas storage during the summer in order to purchase gas when prices are at a seasonal low and utilize it during the winter season when prices tend to be higher[44].

Due to the inability to forecast weather conditions several months in advance, industry participants rely on historical estimates to determine the optimal amount of gas to store and the appropriate timing and conditions for utilization. In the event of a colder than average winter, gas storage may be depleted quicker than expected, resulting in a potential increase in gas prices for the remainder of the season.

Storage facilities are utilized not only as a mechanism to mitigate the seasonal fluctuations in gas prices due to temperature changes, but also as a means of ensuring the continuity of power supply during emergencies. These emergencies can be pipeline disruptions, disagreements between buyers and sellers, or to compensate for issues with other energy sources.

2.5.3 Maintenance

Gas pipes, as with all physical objects, get worn down with time and need maintenance. To perform maintenance, the gas pipe must be shut down, which would decrease the supply [2]. This does not cause a big price surge since these routine maintenance breaks are warned a long time in advance so players can prepare, e.g., filling up storages. Unforeseen accidents however, and the following repair, would not have a warning and therefore cause an immediate price spike. Unlike oil, gas is difficult to transport, and is dependent on physical pipes which leads to regional variations. Therefore it is not easy to substitute gas transport.

2.5.4 EU Emissions Trading System

To combat climate change, the EU initiated the world's first major carbon market in 2005 called the EU Emissions Trading System (ETS)[4]. It functions as a "cap and trade"-scheme by giving the major industries-emitters a quota on how many units of emissions they can emit each period. Industries can trade these quotas in accordance with their pollution needs, which incentives a cleaner operation. EU then reduces the total cap of quotas after each period, effectively making emissions more costly. The direction the ETS influences natural gas prices is unclear due to two opposing effects [11]. Since the ETS penalizes carbon-dioxide emission, and gas consumption produces emission, an increase in the ETS price should therefore negatively effect the gas price. On the other side, power generation from natural gas emit significant less carbon-dioxide compared to power generation from coal, an increase in the ETS should lead to a positive increase in the gas price.

2.5.5 Substitutes for natural gas

Natural gas and coal are both utilized in power plants for electricity generation. The decision on which fuel to use is based on various factors, but predominantly the price. This has resulted in the term Coal-to-Gas switching price, which describes the relative price difference for the most competitive commodity[35]. The switching price is largely based on the difference in price between the two commodities. However it is also based on the difference in emissions between the two fuels, since the carbon pricing must be factored in. Additionally, switching between the two fuels has its own cost.

The correlation between natural gas prices and oil prices is a well-established phenomenon in the energy markets[19]. This means that changes in the price of oil tend to affect the price of natural gas, and vice versa. However, in recent years, there has been a trend towards a delinking of the two commodities, as natural gas prices have become increasingly influenced by their own supply and demand dynamics. This is due to the growing use of natural gas as a fuel source, particularly in power generation, where it has become a alternative to coal. As a result, there is still a correlation between natural gas and oil prices, however there are now more factors at play that can cause natural gas prices to move independently of oil prices.

As concerns over climate change and greenhouse gas emissions have increased, there has been a push towards using greener energy sources such as solar, wind, and nuclear power. While natural gas is a cleaner fossil fuel than coal or oil, it still emits carbon dioxide and other greenhouse gases when burned. Therefore, in the long run, natural gas will likely be substituted by greener energy sources as technology improves and becomes more cost-effective. However, natural gas is still an important energy source today, and it can play a vital role in the transition to greener energy. Natural gas-fired power plants can serve as a reliable backup for intermittent renewable energy sources such as solar and wind power. Additionally, natural gas can be used as a feedstock for the production of hydrogen, which can be used as a zero-emission fuel for transportation and industry. In short, while the ultimate goal is to phase out fossil fuels altogether, natural gas will likely continue to play an important role in the energy mix for some time to come.

3   Data Analysis

The daily closing gas-price data was from ICE Futures Europe and counts 3508 data points ranging from 1/1/2009 to 23/9/2022. This chapter aims to understand the characteristics of the gas-price and for the covariates-data.
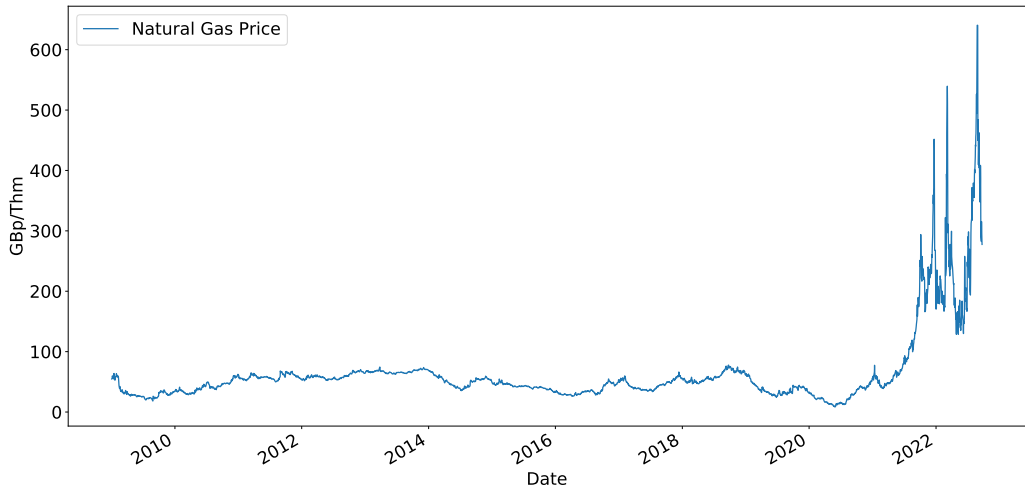
Figure 3: The gas price between 2009-2023.

(a) The natural gas price 2009-2021

(b) The natural gas price 2021-2023
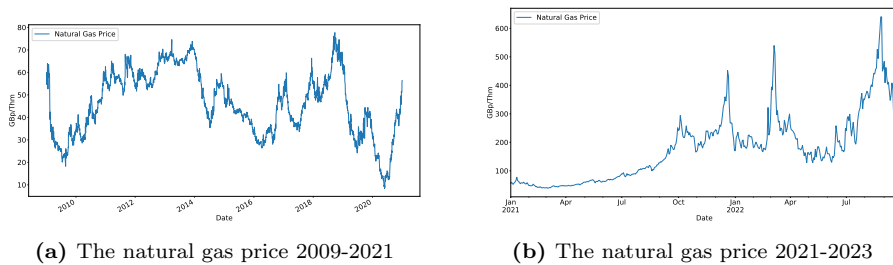
Figure 4

Examining the plots of the gas price in Figure 4, it becomes evident that a distinguishable change in the price dynamics occurred during the year 2021. Notably, the maximum gas price throughout the period from 2009 to 2021 did not exceed 80 GBp/kTherm. However, during the autumn of 2022, the gas price experienced a significant upsurge, reaching the 600 GBp/kTherm threshold at its peak.

To better understand the seasonal influence on the gas price, one can do a seasonal decomposition with the naive model

$$GasPrice(t) = Trend(t) \cdot Seasonalcomponent(t) \cdot Residual(t)$$

The plots in Figure 5 show that there was no significant correlation between the season and price.
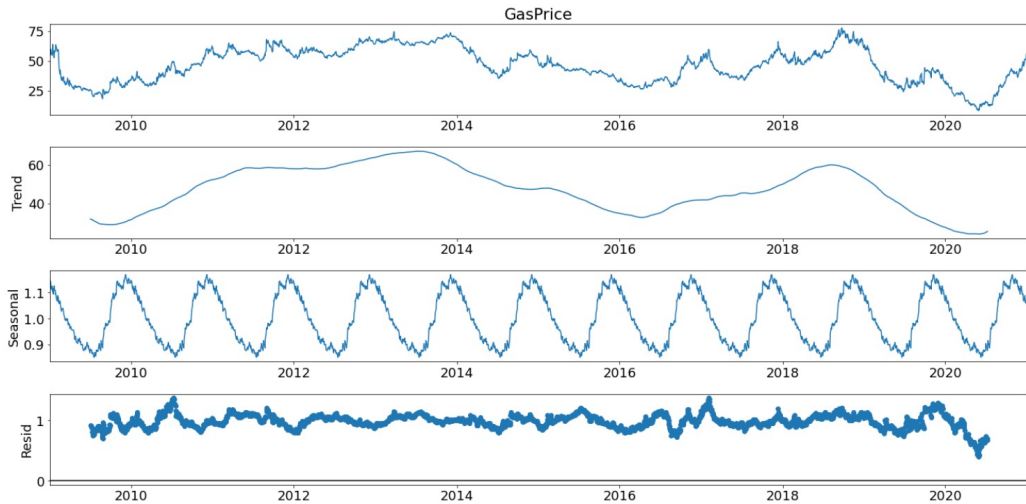
Figure 5: The Seasonal decomposition between 2009-2021.

3.1 Covariates

There are several covariates used, and they are listed in Table 1. The table includes the name of the variables, a description of what they represent, the unit and where they were found.

Examining the variables, shown in Figure 6, it becomes apparent that the coal price experienced a comparable surge in price to natural gas, whereas the oil price did not. Additionally, the price of ETS permits increased rapidly the last few years due to an increase in the periodic withdrawal. Many of these variables displayed a seasonal pattern.

3.2 Preprocessing

The oil price data had 7 missing points, the coal price data had 14 missing points, and the ETS data had 56 missing points. These missing values were distributed evenly across the data set and had been filled using linear interpolation. Notably the variables EU gas storage capacity, gas injection to EU storages and LNG used, started at 3/1/2012, whereas the SOLAR data started at 2/01/2013.

There was one significant downward spike in 2020 for both the Henry HUB price and the oil price. It is possible to interpolate the spikes to smooth the data, however there is valuable insight of how such events can affect the models. Generally, it is a sub-optimal practice to alter actual data.

Table 1: Table with the covariates name, description, unit and source.

Variable	Description	Unit	Source
GasPrice	The natural gas price on the ICE	GBp/Thm	ICE Futures
CoalPrice	Newcastle Coal Futures	USD	Investing.com [23]
OilPrice	Brent Oil Futures	USD	Investing.com
LNG_ Out	Aggregated gas flow out of LNG facility	GWh/day	GIE ALSI [18]
EU_Storage_Cap	Capacity for the aggregated EU gas storages	TWh	GIE AGSI [17]
EU_Inject	Capacity to the aggregated EU storages	TWh	GIE AGSI
EUTrend	The trend of in and out	%	GIE AGSI
nat_ dem	National demand of electricity in the UK	MW/day	nationalgridESO [31]
COAL	Generation from gas	MW/day	nationalgridESO
GAS	Generation from gas	MW/day	nationalgridESO
CWV	Composite Weather Variable	-	meteostat [14]
ETS	Amount of LNG going into the system	EUR/ton CO_2	EEA [15]
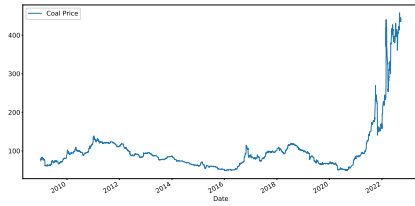
(a) The Coal Price
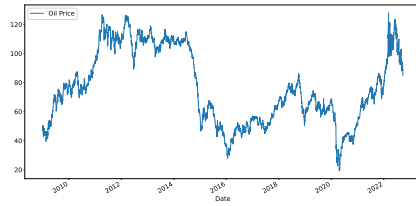
(b) The Oil Price
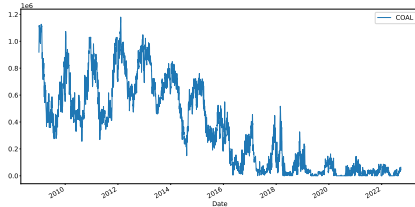
(c) Generation from coal
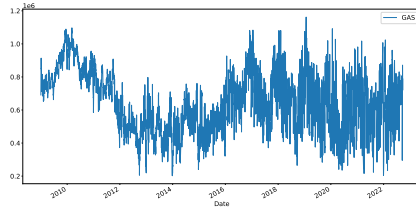
(d) Generation from gas

(e) The aggreagated EU Storage Capacity
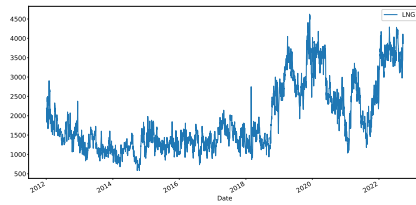
(f) Gas flow out of the LNG facility in the UK
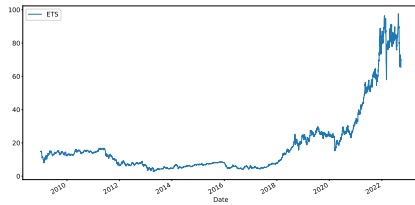
(g) The ETS
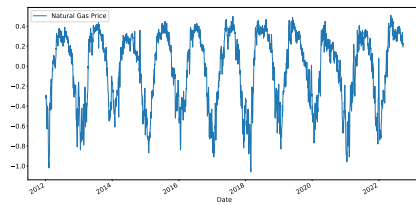
(h) EU gas storage trend
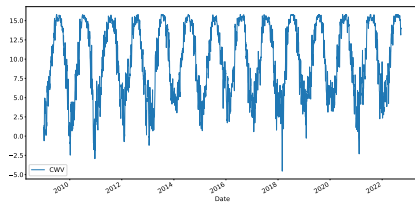
(i) The CWV
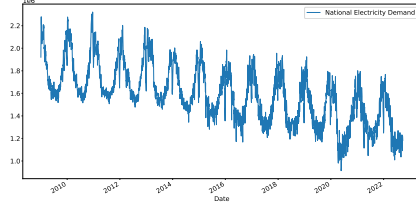
(j) National electricity consumption
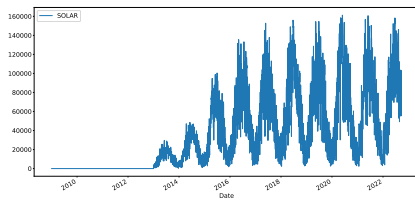
(k) Generation from solar energy
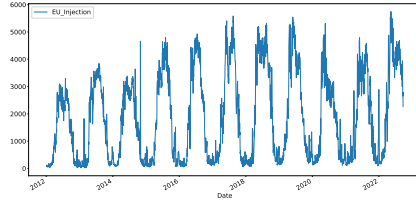
(l) Injection to EU Storages

Figure 6: The variables used in the models
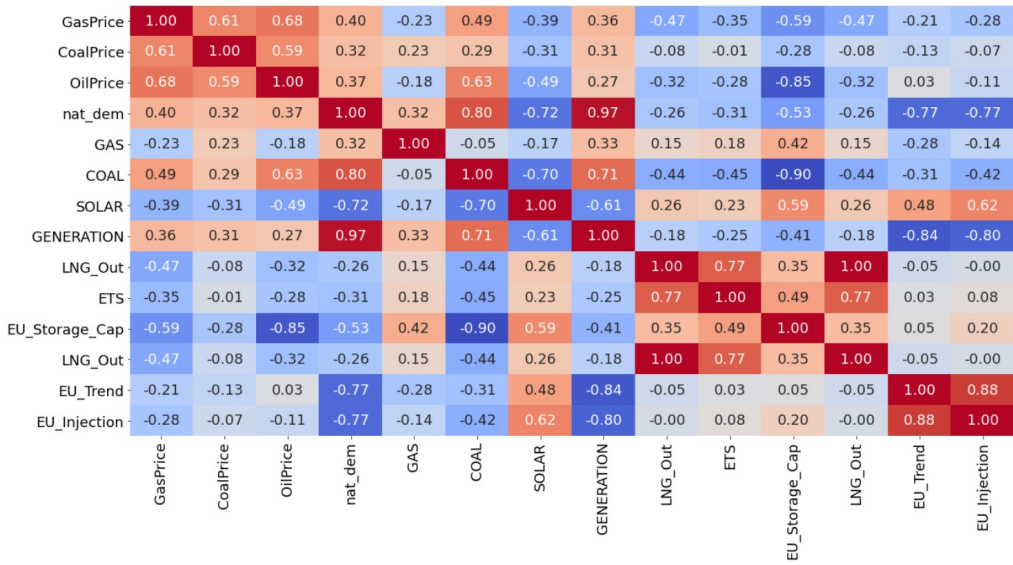
Figure 7: Heat map correlation matrix of the different variables. A stronger color of red is a higher positive correlation, while a stronger color of blue is a stronger negative correlation.

Table 2: Part 1 of the Covariate Data descriptive statistics. The 25%, 50% and 75% refers to quantiles.

Variable	GasPrice	CoalPrice	OilPrice	OilPrice2	CoalPrice2	nat_dem
count	3508	3508	3508	3508	3508	3508
mean	63	102	76	6497	14913	1594671
std	63	67	26	4079	30107	251620
min	8	48	19	374	2352	913771
25%	37	68	56	3085	4610	1406283
50%	49	86	72	5204	7422	1591582
75%	62	105	104	10858	11017	1757626
max	640	458	128	16379	209581	2321378

Table 3: Part 2 of the Covariate Data descriptive statistics. The 25%, 50% and 75% refers to quantiles.

Variable	CWV	GAS	COAL	SOLAR	GENERATION	LNG_Out
count	3508	3508	3508	3508	3508	2750
mean	10	650224	350641	36792	1742475	1917
std	5	182267	317511	40623	215937	882
min	-5	202094	0	0	1195732	583
25%	6	504160	40310	0	1585570	1264
50%	10	657260	304954	21106	1704355	1580
75%	15	783922	615726	63522	1889100	2489
max	16	1162164	1180263	161277	2435884	4625

Table 4: Part 3 of the Covariate Data descriptive statistics. The 25%, 50% and 75% refers to their respective quantiles.

	ETS	EU_Trend	EU_Inject	EU_Storage_Cap
count	3508	2750	2750	2750
mean	19	-0	1872	1040
std	20	0	1587	113
min	3	-1	18	707
25%	6	-0	311	975
50%	12	0	1606	1108
75%	23	0	3273	1116
max	98	1	5749	1135

3.3 Composite Weather Variable

Temperature explains most of the variation in the gas demand, but a better fit can be obtained by including other weather variables [30]. The combination of temperature and other weather variables is called the Composite Weather(CW)[16]. The CW takes into account not only temperature, but also wind speed, effective temperature and seasonal normal effective temperature and is defined as

$$CW = I_1 ET_t + (1 - I_1)\ SNET_t\ - I_2\ W\ \max(0, T_0 - T). \tag{1}$$

Table 5: Table of the variables and their description for the CW.

Variable	Description
T	The temperature
$ET_t = (ET_{t-1} + T)/2$	The effective temperature for day t
T_0	The wind chill temperature cut-off
I_1	The effective temperature weight
I_2	The wind chill weight
W	The wind speed
$SNET_t$	Seasonal normal effective temperature for day t

To model gas demand as a linear relationship to the weather, one can use the CW to define the Composite Weather Variable(CWV) such that,

$$CWV = \begin{cases} V_1 + q(V_2 - V_1), & CW \geq V_2 \\ V_1 + CW(V_2 - V_1), & V_1 < CW < V_2 \\ CW, & V_0 \leq CW \leq V_1 \\ CW + I_3(CW - V_0), & CW < V_0 \end{cases} \tag{2}$$

where the constants are described in Table 6.

The parameters in Figure 8 were empirically calculated by nationalgrid[30] with regards to the effect weather has on the natural gas demand for each LDZ.

For each LDZ, the coordinates to the respective weather stations were used to get weather

Table 6: Table of the constants and their description for the CWV.

constant	description
q	the slope relating to warm weather cut-off
V_0	the cold weather upturn threshold
V_1	the lower warm weather cut-off
V_2	the upper warm weather cut-off
I_3	the cold weather sensitivity

LDZ Weights	0.09	0.10	0.07	0.04	0.13	0.13	0.08	0.12	0.07	0.06	0.09	0.01	0.03	
Parameters	EA	EM	NE	NO	NT	NW	SC	SE	SO	SW	WM	WN	WS	Weighted
I_1	0.7190	0.6910	0.6760	0.6630	0.7270	0.6970	0.6350	0.7120	0.7200	0.6820	0.7200	0.6970	0.6690	0.6982
I_2	0.0144	0.0144	0.0159	0.0086	0.0151	0.0149	0.0119	0.0140	0.0134	0.0100	0.0111	0.0149	0.0101	0.0134
I_3	0.0900	0.0500	0.0000	0.1500	0.2200	0.3000	0.1500	0.3300	0.2400	0.2200	0.1400	0.3000	0.1100	0.1838
V_0	3.0000	3.0000	3.0000	3.0000	3.0000	3.0000	3.0000	3.0000	3.0000	3.0000	3.0000	3.0000	3.0000	2.8037
V_1	15.3000	13.5000	14.7000	13.0000	15.2000	14.9000	12.2000	15.1000	14.8000	14.2000	13.7000	14.9000	14.8000	14.3878
V_2	19.2000	16.8000	17.9000	16.0000	19.2000	18.0000	16.0000	18.7000	18.2000	17.3000	17.2000	18.0000	17.9000	17.8660
q	0.3400	0.4900	0.3800	0.4600	0.3800	0.3800	0.6400	0.3800	0.3700	0.4200	0.4300	0.3800	0.4600	0.4200
W_0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
T_0	14.0000	14.0000	14.0000	14.0000	14.0000	14.0000	14.0000	14.0000	14.0000	14.0000	14.0000	14.0000	14.0000	14.0000

Figure 8: The LDZ constant parameters.

data from metrostat.net python library. For each set of coordinates, the CWV was calculated using the parameter in Figure 8 and afterwards the weighted averages with the individual weights in Figure 9 were used.

LDZ	Weather Station	Weight
SC	Bishopton	9.3
NO	Albermarle	5.5
NW	Rostherne2	12.1
NE	Nottingham_Watnall	6.4
EM	Nottingham_Watnall	10.1
WM	Winterbourne2 (temp), Coleshill (wind)	8.9
WN	Rostherne2	1.1
WS	St.Athan	3.4
EA	Heathrow	8.3
NT	Heathrow	11.4
SE	Heathrow	10.6
SO	Southampton_Oceanographic_Centre	7.1
SW	Filton	5.8

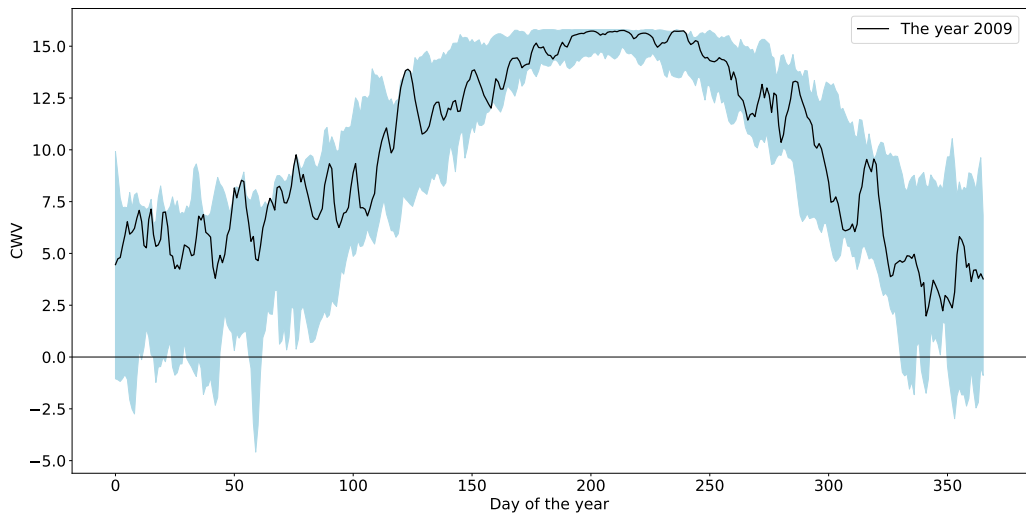**Figure 9:** The LDZ demand weights.

Figure 10: The minimum and maximum values for the CWV from 1990 to 2023 with the CWV for the year 2009 as an example.

4 Modelling Financial Time series

Financial time series, such as stocks, currency-exchange, and commodity prices, can be generalized to share some of the same characteristics [34]. This includes non-stationarity, smooth mean behavior during long time periods and stochastic fluctuation at shorter time-intervals. When trying to model a financial time series, for example the natural gas price, one must focus on capturing the deterministic dynamics of the time series, and accept the irreducible error given by the unpredictable nature of the stochastic fluctuation.

4.1 The value in modelling

The modeling of financial time series holds significant value in the field of finance, providing researchers, practitioners, and policymakers with insights that contribute to decision-making, risk management, and a deeper understanding of the relationship between economic factors and financial markets.

Financial time series modeling enables the prediction of future values based on historical data. Accurate predictions are crucial for informed decision-making, risk assessment, and strategic planning. By analyzing past patterns, trends, and seasonality, time series models allow for forecasting, helping investors, traders, and financial institutions in making well-informed decisions.

Financial markets are strongly influenced by various economic factors, including interest rates, inflation, GDP growth, and geopolitical events. Modeling financial time series helps to understand the intricate relationships between these economic factors and financial market movements. By incorporating economic variables into the models, researchers gain insights into how changes in economic conditions affect asset prices, market volatility, and investor behavior.

Financial markets and economies are subject to constant evolution. Modeling financial time series acknowledge the dynamic nature of the world, where the truth of yesterday may not necessarily hold true today. Time series models provide a framework to capture changing market dynamics, adapt to new information, and update predictions accordingly. They allow for the incorporation of the most recent data, ensuring that the models remain relevant and reliable in an ever-changing financial landscape.

4.2 The requirement of a developed market

The requirement of a developed market for modeling financial time series stems from various critical aspects inherent in such a market[6]. A developed market has a set of attributes, characteristics, and functionalities that play a crucial role in facilitating modeling of financial data. Such a market ensures that there are no "invisible forces" at play, such as undisclosed information or manipulation through insider trading. These practices can distort the underlying patterns and dynamics of financial data, making it challenging to develop reliable and accurate models.

In a developed market, prices are assumed to reflect all available information and quickly adjust to new information[24]. This property, known as price efficiency, is essential for

modeling financial time series. Efficient prices help to identify patterns, trends, and relationships within the data, enabling the construction of reliable models that capture the underlying dynamics of the market.

4.3 Supervised learning

Supervised learning is a well known approach that involves the training of predictive models using labeled data[22]. It is characterized by the presence of a clearly defined target variable or outcome that the model aims to predict based on a set of input features. The process entails utilizing a training data set, comprising input-output pairs, to guide the model in learning the underlying patterns and relationships between the inputs and the corresponding outputs. Through iterative optimization techniques, the model aims to minimize the discrepancy between its predicted outputs and the actual observed values, thus refining its predictive performance.

A crucial step in supervised learning is the division of the available data into training and test sets. The training set is used to train the model, while the test set remains unseen during the training process and serves as an independent evaluation measure of the model's performance. This separation allows for an estimation of the model's ability to generalize to new, unseen data and provides valuable insights into its predictive accuracy.

Overfitting is a common challenge encountered in supervised learning. It occurs when a model becomes excessively complex, capturing noise and distinctive attributes specific to the training data, thereby hindering its generalization to unseen instances. To avoid overfitting, techniques such as regularization are employed to impose constraints on the model's complexity and prevent it from memorizing the training set.

Supervised learning encompasses various algorithms such as decision trees, support vector machines, random forests, and neural networks. These algorithms differ in their underlying mathematical principles and learning mechanisms, but they all share the common objective of inferring a mapping function from the input features to the target variable.

The effectiveness of supervised learning relies heavily on the availability of high-quality labeled data, as it serves as the foundation for model training and evaluation. A well-curated training data set facilitates the learning process, enabling the model to generalize and make accurate predictions on unseen data. Furthermore, the selection and engineering of informative features play a critical role in enhancing the model's performance.

Supervised learning finds applications across various domains, such as classification, regression, and time series analysis. It empowers researchers and practitioners to address a multitude of real-world challenges, ranging from image recognition and natural language processing to financial forecasting and medical diagnosis. By leveraging the power of labeled data, supervised learning algorithms contribute to the advancement of knowledge and the development of intelligent systems with practical utility.
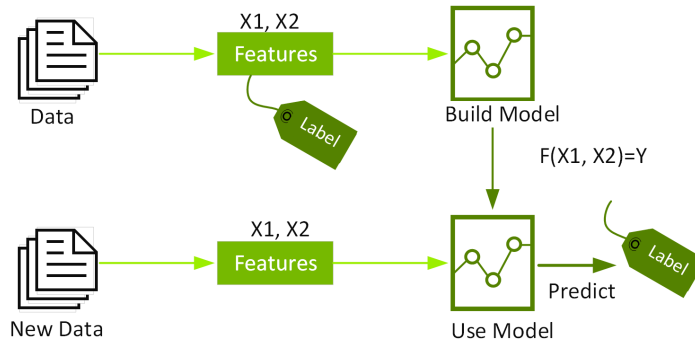
Figure 11: Supervised learning.

4.4   Methods for modelling

There are several ways to model financial time series. Each method has its strengths and weaknesses, and the choice of model depends on the modeling's purpose. A selection of common types of models are listed below:

- **Classical statistical models:** A key advantage of using classical statistical models such as ARIMA or regression models is that the models estimates coefficients which describe the weighting of different covariates or the nature of reacting to previous lags.

- **Machine Learning Methods:** Machine Learning (ML) is a method of building models that try to minimize error with an algorithm. In cases where it can be difficult or unfeasible to use conventional algorithms, ML can be a better alternative.

- **Artificial neural networks:** Neural networks are heavily inspired by biology with regards to the collection of connected units called nodes, which loosely models the neurons in a biological brain[38]. These nodes have weights trained to model a certain variable.

4.5   Model Evaluation

4.5.1   Error metrics

The evaluation of predictive models necessitates the utilization of appropriate error metrics to quantify the discrepancies between predicted and observed values. Different error metrics provide distinct perspectives on model performance, enabling a comprehensive assessment of the predictive accuracy and robustness. Among these metrics, two widely employed measures are the Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE)[10].

The MAE calculates the average absolute difference between the predicted and actual values. Mathematically it is defined as

$$MAE = \frac{\sum_i^n |\hat{y}_i - y_i|}{n} = \frac{\sum_i^n |e_i|}{n},$$

where $\hat{y}_i$ is the predicted value, y_i is the actual value and n is the number of predictions. It offers a straightforward representation of the average magnitude of errors, disregarding the directionality. By summing the absolute differences and dividing by the total number of instances, MAE provides an intuitive measure of the model's average absolute deviation from the true values. MAE is particularly useful when the magnitude of errors holds significance and the directionality of deviations is less crucial.

The RMSE extends the concept of MAE by incorporating a squared error term, yielding a measure that is more sensitive to larger deviations. Mathematically it is defined as

$$RMSE = \sqrt{\frac{\sum_i^n (\hat{y}_i - y_i)^2}{n}} = \sqrt{\frac{\sum_i^n e_i^2}{n}}.$$

RMSE calculates the square root of the average of the squared differences between the predicted and actual values. By squaring the errors, RMSE amplifies the impact of larger errors while maintaining the positive nature of deviations. RMSE is particularly valuable when there is a need to penalize larger errors more severely.

4.5.2   Q-Q plots

Q-Q plots, short for quantile-quantile plots, are graphical tools extensively employed in statistical analysis to assess the degree of similarity between the distribution of observed data and a theoretical distribution, which is typically assumed to be a normal distribution[45]. These plots enable researchers to visually inspect the goodness-of-fit between the empirical quantiles of the observed data and the corresponding quantiles of the assumed theoretical distribution.

To construct a Q-Q plot, the observed data is first sorted in ascending order, and the corresponding quantiles are calculated. Theoretical quantiles are then computed based on the assumed distribution. These quantiles are plotted against each other on a scatter plot, with the observed quantiles on the vertical axis and the theoretical quantiles on the horizontal axis.

By examining the resulting scatter plot, one can assess the agreement between the observed data and the assumed distribution. If the observed data closely adheres to the theoretical distribution, the points in the Q-Q plot will lie approximately on a straight line with a 45 degree angle. Deviations from the line suggest that the data may not follow the assumed distribution.

Q-Q plots are particularly useful to identify departures from normality, as they highlight systematic deviations in the tails or center of the distribution. Skewness, heavy tails, or other departures from normality become apparent through the nonlinear patterns exhibited in the Q-Q plot. Deviations from linearity can indicate potential issues such as outliers, heteroscedasticity, or the need for alternative distributional assumptions.

5 Multiple Regression

Multiple linear regression is a statistical method to better understand the linear correlation between a single dependent variable and two or more independent variables. By using previously known data, one can assign weights to the independent variables based on their relative contribution to an estimate of the dependent variable. In mathematical notation this is written as

$$\hat{y} = w_0 + \sum_{j=1}^{k} w_j x_j + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2) \tag{3}$$

where w_j being the coefficients of the parameter, c is a constant and the x_j's are the variables and $\hat{y}$ is the response variable. The equation can also be written as

$$\hat{y} - w_0 = \sum_{j=1}^{k} w_j (x_j - \bar{x}_j) + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$$

with $\bar{x}_j$ being the average of the j-th covariate over the n observations.

5.1 Parameter estimation

The coefficients w_{ik} for a set of points $\{x_{i1}, ...x_{ik}\}$ are obtained by minimizing the loss function, defined as

$$\sum_{i=1}^{n} (\hat{y}_i - y_i)^2. \tag{4}$$

The $\hat{y}_i$ is the models best prediction for the i-th observation in a set of n observations[48].

$$\sum_{i=1}^{n} (\hat{y}_i - y_i)^2 = \sum_{i=1}^{n} \left(w_0 + \sum_{j=1}^{k} w_j (x_{ij} - \bar{x}_j) - y_i \right)^2$$

The right hand side can also be viewed as a function of the w_i, given any values of the pairs $(x_{11}, ..., x_{1k}, \ y_1), \ ... \ , (x_{n1}, ..., x_{nk}, \ y_n)$. This new function is defined as

$$g(w_0, ...w_k) = \sum_{i=1}^{n} \left(w_0 + \sum_{j=1}^{k} w_j (x_{ij} - \bar{x}_j) - y_i \right)^2$$

In order to find the minimum of the expression, one must calculate its derivatives and set them equal to zero.

$$\frac{\partial g}{\partial w_j} = 2\sum_{i=1}^{n}(x_j - \bar{x}_j)\left(w_0 + \sum_{m=1}^{k} w_m(x_{im} - \bar{x}_m) - y_i\right) = 0$$

$$\frac{\partial g}{\partial w_0} = 2\sum_{i=1}^{n}\left(w_0 + \sum_{m=1}^{k} w_j(x_{im} - \bar{x}_m) - y_i\right) = 0$$

Note the changing of indexing from j to m. Simplifying these equations one get

$$\sum_{i=1}^{n}(y_i - w_0)(x_{ij} - \bar{x}_j) = \sum_{i=1}^{n} w_i \sum_{m=1}^{k} w_j(x_{im} - \bar{x}_m)(x_{ij} - \bar{x}_j)$$

$$\sum_{i=1}^{n} y_i = n \cdot w_0 + \sum_{i=1}^{n}\sum_{m=1}^{k} w_j(x_{im} - \bar{x}_m)$$

Knowing that $\sum_{i=1}^{n}(x_{ij} - \bar{x}_j) = 0$, the last equation can be reduced to

$$w_0 = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{y} \tag{5}$$

The other equations are

$$\sum_{i=1}^{n}(y_i - w_0)(x_{ij} - \bar{x}_j) = \sum_{i=1}^{n} w_i \sum_{m=1}^{k} w_j(x_{im} - \bar{x}_m)(x_{ij} - \bar{x}_j)$$

which is the same as writing

$$\mathrm{cov}(y, x_j) = \sum_{m=1}^{k} w_m \mathrm{cov}(x_m, x_j). \tag{6}$$

Since there are k unknowns with k equations there is a solution for the w_m's.

5.2 The problem of multicollinearity

A key part of multiple regression is to understand the linear relationships between each covariate and the response variable[5]. The goal is to achieve a relationship where an unit increase in a single covariate, the models prediction would be changed by a magnitude of the belonging coefficient. This becomes harder when the covariates are correlated, since a change in one of them can change the values for the other covariates. If this is the case, the regression coefficient would become unstable and thereby change significantly with different data-samples. This would reduce the statistical significance of a covariate. This problem is known as multicollinearity. It is worth noticing that the predictive power of the regression model is not affected by multicollinearity.

5.3 Variance Inflation Factor

Variance Inflation Factor (VIF) can be used to check for multicolliniarity between covariates[47]. First, a multiple regression for the m-th covariate is estimated by the equation

$$\hat{x}_m = w_0 + \sum_{j \neq m}^{k} w_j x_j$$

with k covariates. Using the definition of R^2, the coefficient of determination for the m-th regression model is denoted

$$R_m^2 = 1 - \frac{\sum_i^n (x_{im} - \hat{x}_{im})}{\sum_i^n (x_{im} - \bar{x}_m)}.$$

Finally, the VIF is calculated by

$$\text{VIF}_m = \frac{1}{1 - R_m^2}.$$

An independent covariate, i.e., no multicollinearity, has a score of 1. In general, a VIF higher than 4 indicates some multicollinearity, while a score higher than 10 indicates significant multicollinearity.

5.4 Weighted observations

When dealing with time series data that exhibits temporal changes, utilizing a regression model proves suboptimal as it tries to attain an optimal fit for the entire dataset. This is problematic since the first time points can be very different from the present time points, effectively hurting the future predictions. To address this issue, it is possible to assign greater significance to recent observations by weighting the data with a bias towards recent time-points. One approach to accomplish this, is to introduce an auxiliary variable function W defined as

$$W(t) = \left(\frac{t}{n} \right)^{\lambda}$$

where λ is a scaling constant and n is the number of data points. For time series data, the index i denoting the observations, is changed to t to highlight time steps. A higher value of λ gives higher importance to the more recent data points. Subsequently, this auxiliary variable function can be integrated into the loss function in equation 4 to generate a revised loss function

$$\sum_{t=1}^{n} \frac{(\hat{y}_t - y_t)^2}{W(t)}. \tag{7}$$

Algorithm 1 The MLR algorithm

Input : Train and test data, weight-constant λ

Define a loss function
Use Train to Solve the k equations in 6 to get the coefficients

Use coefficients to predict Test

Output: $\hat{y}$

6 Gradient Boosting

Gradient Boosting is a statistical technique employed for developing prediction models geared towards regression and classification tasks[40]. The fundamental principle underlying gradient boosting involves constructing an ensemble of weak prediction models, commonly in the form of trees, through a sequential process, whereby each model is optimized to minimize the errors of its predecessor. The integration of these models lead to the formation of a strong learner. Notably, the term "gradient" refers to the ability to utilize any differentiable loss function when evaluating errors.

6.1 Ensemble learning

An ensemble learning algorithm refers to the combination of several machine learning algorithms to enhance the overall efficacy of the model. Various ensemble algorithms utilize multiple decision trees to construct a comprehensive model. A decision tree is an algorithm that poses multiple binary if/else questions in a tree-like structure with branches and computes the minimum number of branches required to achieve the best possible outcome. An instance of an ensemble learning algorithm that utilizes decision trees is the Gradient Boosting Decision Trees (GBDT)[8]. This decision tree ensemble learning algorithm shares many similarities with Random Forest, although with notable distinctions.

Random Forest and GBDT share a similar model structure that involves multiple decision trees. However, the difference between the two techniques lies in the methodology of the tree construction and integration. Random Forest utilizes a technique referred to as bagging to construct complete decision trees in parallel from random bootstrap samples of the data set. The ultimate prediction is derived by averaging the predictions of all decision trees. In contrast, GBDT adopts a sequential approach to build decision trees, resulting in a tree ensemble that is generated iteratively.
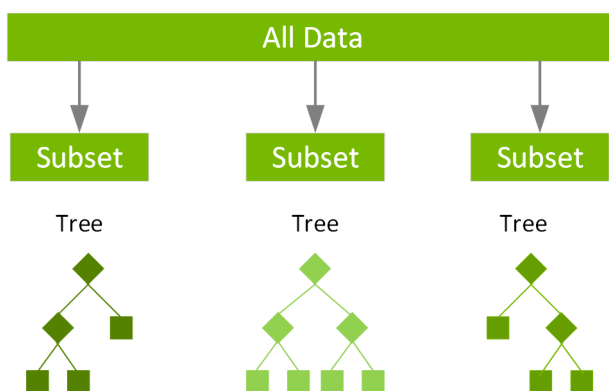

Figure 12: The structure of the tree-building process for Random Forest.

Source: [32]

6.2 Extreme Gradient Boosting

A popular application of gradient boosting is the Extreme Gradient Boosting(XGBoost) machine learning method [12]. Due to some favorable techniques, it has become a favorite on the machine learning platform Kaggle and is considered as one of the current best ML-method to perform regression[12].

A disadvantage with XGBoost, as well as other tree-based methods, is its problem of extrapolating target values that lies outside of the range in the training data[28]. This is because XGBoost partitions the infinite input space into a finite set of possibilities. Consequentially, all tree-based methods cannot model a continuous function with great success.
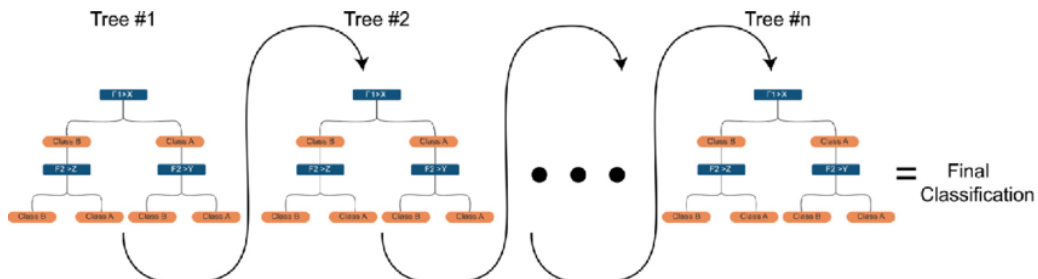
Figure 13: The structure of the sequential tree-building process for XGBoost.

<div align="right">Source: [1]</div>

6.3 The technical details of Extreme Gradient Boosting

6.3.1 The model

The XGBoost model is an ensemble of trees and the model for a single response variable $\hat{y}_i$ can be written mathematically as

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in \mathcal{F}$$

with K being the number of trees, x_i is a vector of features for the i-th data point, f_k being a function(a tree) in the function space $\mathcal{F}$, and $\mathcal{F}$ being the set of all possible trees[7].

$$f_t(x) = w_{q(x)}, \quad w \in R^T, \ q : R \to 1, 2, ..., T,$$

where w is a vector, T is the number of leafs and q connects the data points to the belonging leaf node.

6.3.2 The Training Process

The idea of XGBoost, is to define an objective function and further optimize it[7]. For step t, this objective function is

$$Obj^{(t)} = \sum_{i=1}^{n} L(y_i, \hat{y}_i) + \sum_{i=1}^{t} \Omega(f_i) \tag{8}$$

where the first term is the loss-function, and the second term being the regularization term.

Knowing that XGBoost uses a sequential boosting method, one can write

$$\hat{y}_i^{(t)} = \sum_{k=1}^{K} f_k(x_i) = \hat{y}_i^{(t-1)} + f_K(x_i)$$

which can be substituted into the objective function in equation 8 giving

$$Obj^{(t)} = \sum_{i=1}^{n} L(y_i, \hat{y}_i^{(t-1)} + f_K(x_i)) + \sum_{i=1}^{t} \Omega(f_i).$$

Using the second order Taylor-expansion to quickly optimize the loss-function, the objective function can be approximated by

$$Obj^{(t)} \approx \sum_{i=1}^{n} [L(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_i^2(x_i)] + \sum_{i=1}^{t} \Omega(f_i)$$

where

$$g_i = \frac{\partial L}{\partial \hat{y}_i^{(t-1)}} L(y_i, \hat{y}_i^{(t-1)}), \quad h_i = \frac{\partial^2 L}{\partial(\hat{y}_i^{(t-1)})^2} L(y_i, \hat{y}_i^{(t-1)}).$$

The regularization term

$$\Omega(f_t) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^{T} w_j^2$$

where T is the number of terminal nodes, called leaves, and γ is a user-definable penalty for pruning. The second term is the ridge regularization.

The term $\sum_{i=1}^{n} L(y_i, \hat{y}^{t-1})$ can be omitted when optimizing the objective function due to it being a constant. By substituting in $w_{q(x_i)}$ using equation 8, one arrives at the equation

$$Obj^{(t)} = \sum_{i=1}^{n}[g_i w_{q(x_i)} + \frac{1}{2}h_i w_{q(x_i)}^2] + \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{9}$$

$$= \sum_{t=1}^{T}[G_j w_j + \frac{1}{2}(H_j + \lambda)w_j^2] + \gamma T \tag{10}$$

where $G_j = \sum_{i \in I_j} g_i$, $H_j = \sum_{i \in I_j} h_i$, with the set of indices of data points assigned to the j-th leaf is $I_j = \{i|q(x_i) = j\}$.

In order to find the optimal value w_j^*, the term γT can be omitted. Further, the argmin $Obj^{(t)}$ can be found by solving for each quadratic equation. Knowing that w_j are independent of each other, the optimal value for w_j can be derived by solving

$$G_j w_j + \frac{1}{2}(H_j + \lambda)w_j^2 = 0 \quad \longrightarrow \quad w_j^* = -\frac{G_j}{H_j + \lambda}$$

Substituting in the optimal weight in to equation 10, the minimum objective function is given by

$$\min Obj^{(t)} \approx -\frac{1}{2}\sum_{t=1}^{T}\frac{G_j^2}{H_j + \lambda} + \gamma T. \tag{11}$$

6.3.3 The tree-building

The computational limits make it impossible to compute all possible trees, so the XGBoost iteratively grows trees using the information from the previous trees to minimize the error. The tree building process involves decisions of whether splitting a leaf into two new leaves or not. The decision is based on the potential of overfitting, using the metric called Gain, which is defined as

$$Gain = \frac{1}{2}\left[\frac{G_L^2}{H_L + \gamma} + \frac{G_R^2}{H_R + \gamma} - \frac{(G_L + G_H)^2}{H_L + H_R + \gamma}\right] - \gamma. \tag{12}$$

If $Gain > 0$ the leaf is split. In this equation, the first and second term are the scores of the two potential new leaves, with the third being the original leaves. The last term is the regularization term which punishes extra leaves.

6.3.4 The Algorithm

Algorithm 2 The tree-bulding Algorithm

    Input : $\{(x_i, y_i)\}$, Chosen set of hyperparameters
    Initialize a single leaf tree $f_0(x_i)$
    for t in $(1...T)$ **do**
        Compute G_t and H_t
        while $Gain > 0$ **do**:
            Add leafs to $f_t(x_i)$
        end while
        $\hat{f}_t(x_i) = \hat{f_{t-1}}(x_i) + \alpha \hat{f}_t(x_i)$
    end for

    Output : $\hat{f}(x) = \hat{f}_T(x)$

The algorithm produces the final model $\hat{y_i}$ which is an additive culmination of the previous models

$$\hat{y_i} = \hat{f}_T(x_i) = \sum_{k=0}^{T} \hat{f}_k(x_i) \tag{13}$$

6.3.5 Algorithmic enhancements

XGBoost is a type of GBDT, but has some algorithmic enhancements and an optimized system making it significantly greater than GBDT in general[29]. For instance, it uses both ridge and lasso regularization to prevent overfitting and has an automated cross-validation algorithm in place. While a GBDT split nodes until there are no further improvements in a greedy fashion, XGBoost prunes the tree backwards with the help of the 'max_depth' parameter, yielding better computational performance. Another positive difference is that XGBoost automatically interpolates missing values to reduce training loss.

6.4 Hyperparamters

There are several hyperparameters to tune for XGBoost. Some of the most important are listed below[27]:

- Max Depth: Specifies the maximum depth of each decision tree. Deeper trees can capture more complex relationships but are more prone to overfitting.

- The learning rate, η: Controls the step size shrinkage used to prevent overfitting. A lower learning rate makes the model more robust to overfitting but requires more boosting iterations.

- Min Child Weight: Specifies the minimum sum of instance weights needed in a child node. Higher values can help to reduce overfitting by adding constraints on the tree structure.

- The Regularization Parameters (λ and α): These parameters control L1 and L2 regularization, respectively. They help to prevent overfitting and improve generalization by adding penalty terms to the loss function.

- The number of estimators(trees): Determines the total number of boosting iterations or the number of trees in the ensemble. Increasing the number of trees can improve performance, but it also increases the computational cost.

Different hyperparameters can significantly impact the model's performance. A grid search methodically searches through different combinations of these hyperparameters, iteratively evaluating the model's performance using cross-validation techniques. This systematic exploration helps in mitigating the risk of suboptimal parameter configurations and aids in identifying the ideal set of hyperparameters that maximize the predictive power of the XGBoost model.

6.5   Feature Importance

XGBoost measures feature importance by evaluating "Gain"[7]. Gain, defined mathematically in equation 12, is a measure of the improvement of the accuracy which a single feature contributes with when splitting a leaf into two new leaves. In other words, it is the degree of improvement. Averaging the Gain for each variable, results in a measurement on the relative degree of importance for which each feature contributes with.

7 Artificial Neural Network

An Artificial Neural Network comprises a set of interconnected nodes or units known as artificial neurons, which loosely mimic the neurons present in biological brains[38]. Analogous to synapses in the biological brain, each connection permits the transmission of a signal to other neurons. Upon receiving signals, an artificial neuron processes them and relays signals to the neurons connected to it. The signal at a connection is a real number, while the output of each neuron is determined by a non-linear function of the summation of its inputs. These connections are called edges, and neurons and edges possess weights that alter as the system learns. The weight modification process augments or diminishes the signal strength at a connection. Additionally, neurons may have a threshold that activates only if the total signal surpasses the specified threshold value.

Neurons typically manifest in a layered configuration, with each layer executing discrete transformations on the input data. The information flow starts with the initial layer, designated as the input layer, and propagates through intermediate layers before culminating at the terminal layer, known as the output layer.

7.1 Long Short-Term Memory

Long Short-Term Memory (LSTM) is a type of artificial neural network that finds application in the domains of deep learning and artificial intelligence[43]. Unlike the conventional feed-forward neural networks, LSTM networks feature feedback connections. These connections within a LSTM network qualify it as a type of recurrent neural network (RNN) that can effectively process not only discrete data points such as images, but also entire data sequences such as speech or video. This characteristic gives LSTM networks an unique capability for efficiently processing and predicting data. The term 'Long Short-Term Memory' reflects the primary concept on how the LSTM network is an improved RNN since it incorporates both long-term and short-term memory modules. LSTM networks have been utilized in numerous tasks such as handwriting recognition, speech recognition, machine translation and time series forecasting.

One of the primary challenges faced while training RNNs is the vanishing gradient problem [43]. During backpropagation through time, the gradients decrease exponentially, leading to the RNN updating its weights at a slow pace or not at all. This can result in the network being unable to learn meaningful representations of the data or experiencing difficulties in capturing long-term dependencies in sequential data. In order to circumvent this challenge, the LSTM architecture was introduced, which includes a gating mechanism that controls the flow of information in and out of the memory cell. The LSTM architecture comprises three gates, namely the input gate, forget gate, and output gate, that regulate the flow of information into the memory cell and selectively retain or discard information based on its significance. The inclusion of the memory cell and gating mechanism in LSTM networks allow the model to maintain long-term dependencies in the data while avoiding the vanishing gradient problem.
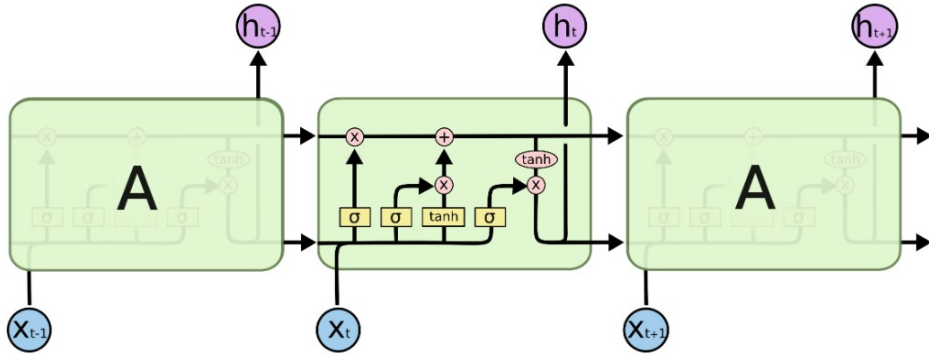
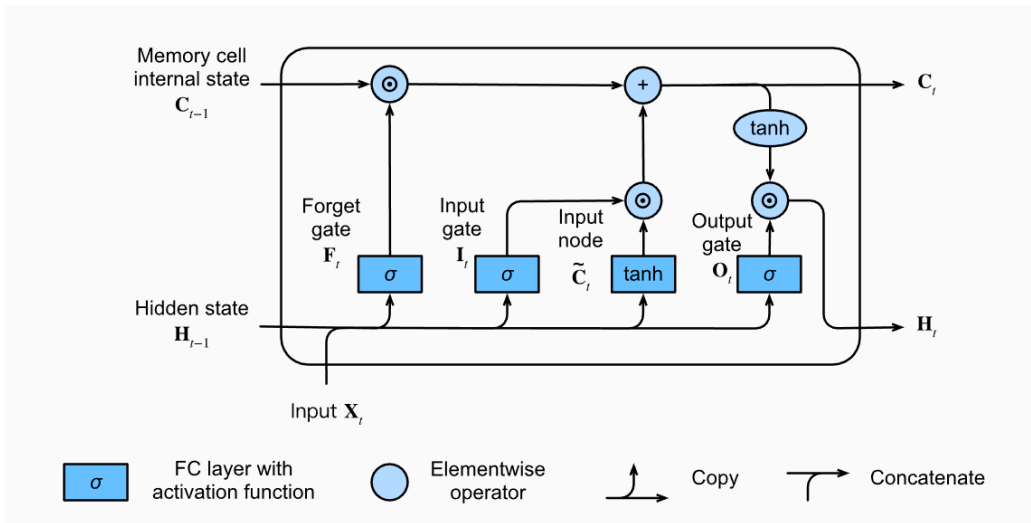**Figure 14:** The figure shows the architecture of the LSTM network.

Figure 15: The figure shows the architecture of the LSTM unit.

7.2 The step-wise procedure of the LSTM

7.2.1 Input and Output

Each LSTM unit receives 3 input vectors, the memory cell internal state $C_{t-1} \in \mathbb{R}^{n \times u}$, called the "long" memory, the hidden state $H_{t-1} \in \mathbb{R}^{n \times u}$, called the "short" memory, and the input vector $X_t \in \mathbb{R}^{n \times d}$ submitted to the LSTM at instant t. The number of hidden units is denoted by u, while the batch size is n and the number of features is denoted by d.

Both vectors C_{t-1} and H_{t-1} are generated by the previous LSTM instant. The two vectors X_t and H_{t-1} combine right away the new inserted information and the recent past information into a concatenated vector $Z_t \in \mathbb{R}^{n \times u+d}$. The concatenated vector Z_t is used to update the previous long-term memory C_{t-1}, which is then used to update the previous short-term memory H_{t-1}. The output of the LSTM-cell are then the hidden state H_t and the cell state C_t

7.2.2 Gates

There are three gates which are used in the LSTM unit to decide what information to pass on. These gates are called the forget-, input- and output gate, and their task is to multiply the incoming vector with weights and add a bias, before using the Sigmoid activation function

$$\sigma(x) = \frac{e^x}{e^x + 1},$$

to produce a vector with values $\in [0, 1]$. The output values represents the element-wise percentage of the information to keep for the incoming vector. The input vector for all three gates is the concatenation of the input vector X_t and the previous hidden state vector H_{t-1}.

7.2.3 The step-wise execution of the unit

The first step of the LSTM unit, shown in figure 15, is to decide what information to keep from the current cell state C_{t-1}. This is done in the forget gate with the function F_t

$$F_t = \sigma(a_F), \quad \text{where} \quad a_F = W_F \cdot Z_t + b_F$$

with weights $W_F \in \mathbb{R}^{d+u \times u}$ and bias $b_F \in \mathbb{R}^u$. The output of F_t is then multiplied element-wise with C_{t-1}.

The second step is to get the selector vector from the input gate with the function I_t

$$I_t = \sigma(a_I), \quad \text{where} \quad a_I = W_I \cdot Z_t + b_I$$

with weights $W_I \in \mathbb{R}^{d+u \times u}$ and bias $b_I \in \mathbb{R}^u$. This vector is then multiplied element-wise with the output vector from the Input node function $\widetilde{C}_t$

$$\widetilde{C}_t = \tanh(a_C) \quad \text{where} \quad a_C = W_C \cdot Z_t + b_C$$

with weights $W_C \in \mathbb{R}^{d+u \times u}$ and bias $b_C \in \mathbb{R}^u$. For this function, instead of the Sigmoid activation function, the hyperbolic tangent function

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

is used as the activation function to ensure that all values are normalized between $[-1, 1]$. The Hademar product $\odot$(element-wise multiplication) of the output from the Input gate

and the Input node is then added on to Memory cell state to complete the update of C_{t-1} to C_t

$$C_t = F_t \odot C_{t-1} + I_t \odot \widetilde{C}_t.$$

The final step is to get the selector vector from the Output gate function O_t

$$O_t = \sigma(a_O), \quad \text{where} \quad a_O = W_O \cdot Z_t + b_O$$

with weights $W_O \in \mathbb{R}^{d+u \times u}$ and bias $b_O \in \mathbb{R}^u$. This vector is then multiplied element-wise with the hyperbolic tangent function on the cell state vector C_t in order to control the stability of the network over time by normalizing the values to be between $[-1, 1]$. The final output of the LSTM unit is then H_t

$$H_t = O_t \cdot \tanh(C_t)$$

7.2.4 Updating weights and biases with backpropagation

Backpropagation is an algorithm in the field of neural networks which is used to train parameterized networks with differentiable nodes, including feedforward artificial neural networks. By applying the chain rule, the algorithm effectively calculates the gradient of a loss function with respect to the weights of the network for a single input-output example. The backpropagation algorithm applies the chain rule with time dependence to compute the gradients, which propagate backward in time through the recurrent connections of the network. Gradient descent and its variants, such as stochastic gradient descent, are frequently employed in combination with backpropagation to optimize the network's weights and biases. Through this process, the LSTM network can learn the optimal values of its parameters and thereby improve its performance on the given task. As such, backpropagation through time represents a key component in the training of LSTM networks, and its effective use can significantly enhance their capabilities.

Using the derivatives given in the Appendix, one can update each of the 4 parameters with the learning rate α after every n time step

$$W_{[F,I,O,\widetilde{C}]} += \alpha \cdot \frac{\partial J}{\partial W_{[F,I,O,\widetilde{C}]}}, \qquad \text{where} \quad \frac{\partial J}{\partial W_{[F,I,O,\widetilde{C}]}} = \sum_{t=1}^{t+n} \frac{\partial J^{(t)}}{\partial W_{[F,I,O,\widetilde{C}]}^{(t)}} \qquad (14)$$

$$b_{[F,I,O,\widetilde{C}]} += \alpha \cdot \frac{\partial J}{\partial b_{[F,I,O,\widetilde{C}]}}, \qquad \text{where} \quad \frac{\partial J}{\partial b_{[F,I,O,\widetilde{C}]}} = \sum_{t}^{t+n} \frac{\partial J^{(t)}}{\partial b_{[F,I,O,\widetilde{C}]}^{(t)}}. \qquad (15)$$

7.2.5 LSTM hyper parameter tuning

Parameter tuning for LSTM models involves careful consideration of key parameters such as the number of units, epochs, and batch size. The number of units in an LSTM layer

determines the complexity and representational capacity of the model. A higher number of units allows the model to capture more intricate patterns in the data but also increases computational demands. It is important to strike a balance between model complexity and computational efficiency.

Epochs refer to the number of times the model iterates over the entire training data set during the training process. Increasing the number of epochs allows the model to learn from the data for a longer duration, potentially improving its predictive performance. However, using too many epochs can lead to overfitting, where the model becomes overly specialized to the training data. Therefore, finding an optimal number of epochs that achieves good generalization is crucial.

Batch size is the number of samples used in each update of the model's weights during training. A larger batch size can improve training efficiency by leveraging parallelism but may also introduce noise and hinder convergence. Conversely, a smaller batch size may offer more accurate weight updates but can slow down the training process. Determining the appropriate batch size involves considering computational resources, convergence speed, and the trade-off between noise and accuracy.

7.2.6   The Algorithm

Epoch, units and batch are three crucial hyperparameters in neural networks. An epoch refers to the number of times the algorithm iterates through the entire dataset, whereas the batch size refers to the number of samples used to work through before updating the internal model parameters. The number of units determines the model's capacity to capture and store information from previous time steps, allowing it to learn complex temporal patterns and dependencies in the data.

Algorithm 3 The LSTM algorithm

Input : Train and test data, n_epochs, n_batch, n_units

Normalize the data
Initialize weights and biases with random numbers
Configured the LSTM network: Activation function, loss function,layers

for epoch in n_epochs **do**
    for batch in n_batch **do**
        Run algorithm on batch
        Update weights and biases with equation 14 and equation 15
    end for
end for

Predict new data with the fitted network
Re-scale the prediction to the original size

Output: $\hat{y}$

8 Results and Discussion

8.1 The training and test data

To evaluate a model, it is beneficial to test on different scenarios. Firstly it is necessary to test a model on a set of "normal conditions" i.e. relatively similar values to the values used in the training. Secondly it is interesting and useful to know how the model handles a new unseen scenario like the Russian Invasion of Ukraine. The models goal was to predict the next 5 days, in other other words 1 trading week, before re-training with the added week as training-data and then repeating the process.

8.2 Model assembly

The tuning of parameters is essential in constructing accurate predictive models, especially when it comes to machine learning and neural networks. This process demands careful attention in order to achieve optimal performance. Determining the optimal parameter values is not straightforward, as different parameters give varying performance. Additionally, it is essential to exercise caution to avoid overfitting, a situation where the model becomes too closely tailored to the training data, resulting in poor generalization to new data. Balancing the trade-off between model complexity and generalization is important.

8.2.1 MLR weighting parameter

The weighting chosen for the two regression models was an exponential scaling of a linear series between 0 and 1 raised to the power of 10. This approach was deemed a favorable compromise as it effectively balanced the objectives of mitigating overfitting and ensuring model robustness, while concurrently addressing the concern of potential ill-conditioned behavior. Ill-conditioned models refer to situations where small changes in the input data lead to significant fluctuations or instability in the output. By utilizing the exponential scaling approach, the model achieved a suitable balance by preserving the underlying linear relationship, minimizing the risk of overfitting, and preventing the data from becoming ill-conditioned, thereby promoting more reliable and stable predictions.

Another important aspect of building a regression model is to check for multicolinearity. By using VIF, the variables of the coal price and the Composite Weather Variable were found to be uncorrelated and thereby suitable for the model.

8.2.2 XGBoost Optuna Grid search

To achieve optimal parameter configurations for XGBoost, Optuna Grid search was used [33]. Optuna Grid search employs a systematic approach to explore the hyperparameter space, allowing for an exhaustive evaluation of various parameter combinations.

Moreover, Optuna Grid search enhances the efficiency of the hyperparameter tuning process by leveraging parallelization techniques, enabling the simultaneous evaluation of mul-

tiple parameter combinations. This parallelized search speeds up the exploration of the hyperparameter space, resulting in faster convergence towards the optimal parameter configuration.

The parameters underwent regular updates at intervals of 4 weeks to optimize their performance as time progressed.

8.2.3 LSTM hyperparameters

The three hyperparameters epoch, units and batch were tuned by performing a grid search with combination of the parameters. The parameters underwent regular updates at intervals of 4 weeks to optimize their performance as time progressed.

8.3 A normal period

The natural gas price is a affected by a multitude of factors and is always changing. However, a certain normal situation could be observed with the minimum and maximum value being constant for a long time period. It was therefore of interest to evaluate the performance of the models during a "normal" period.
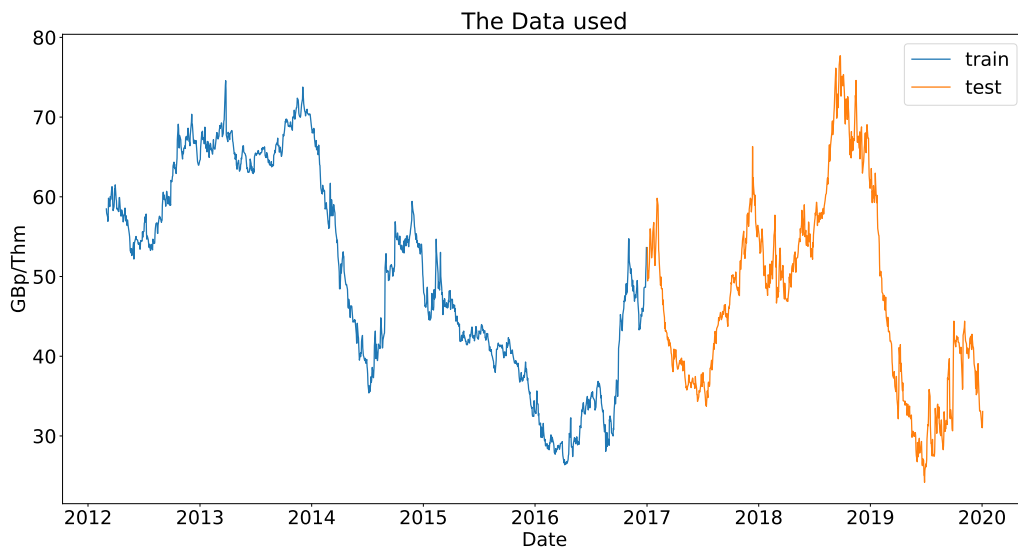

Figure 16: The training and test data for for the normal time period from 1/1/2009 - 6/1/2020.

Examining the predicted gas prices from the three models in Figure 17 with the errors displayed in Table 7, it was evident that the LSTM model provided the most accurate predictions, slightly surpassing XGBoost. The MLR model performed the worst with regards to prediction. Notably, the MLR model struggled more compared to the other two models during the period between the middle of 2018 and 2020. While XGBoost overall captured the price trends, it exhibited occasional downward spikes that were worth noting. All three models demonstrated an accurate understanding of the price dynamics throughout the year 2017.
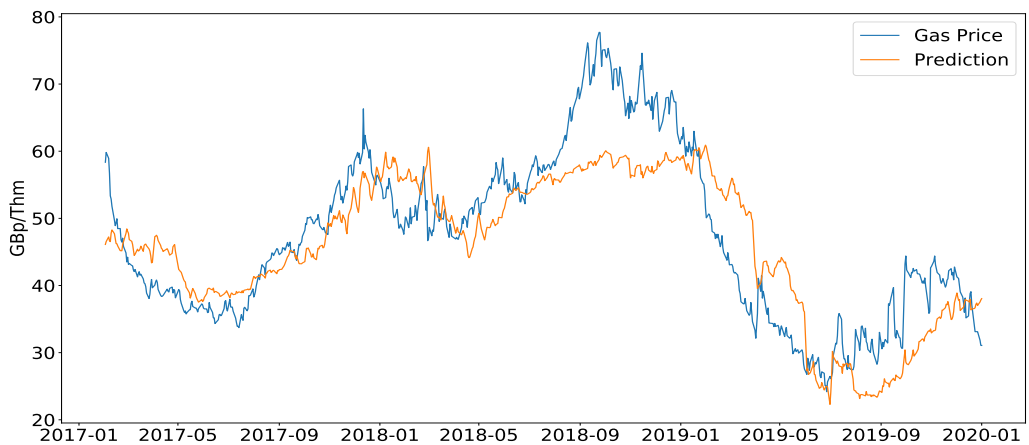
(a) The predicted and actual values for the natural gas price with MLR for the time period 1/1/2017 - 1/6/2020.
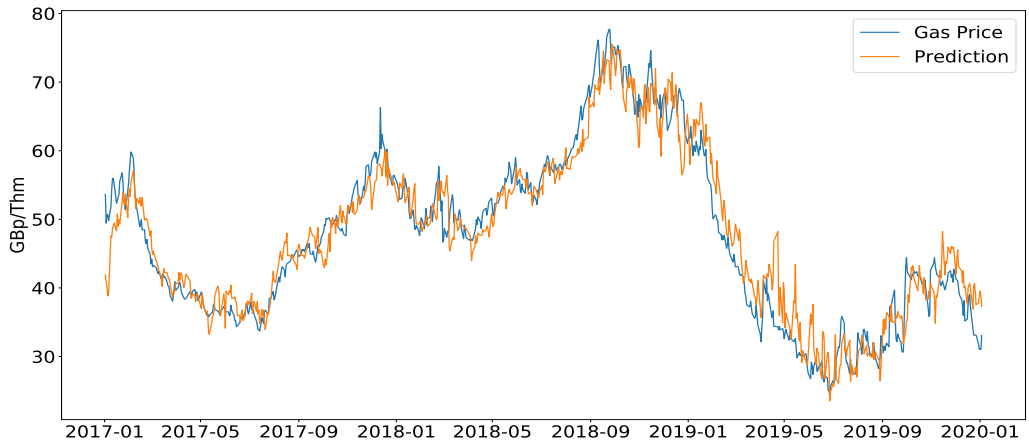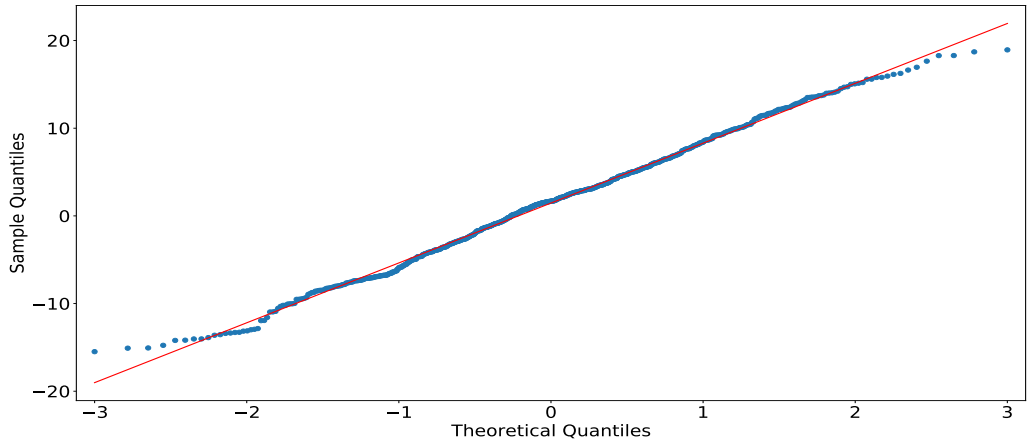

(b) The predicted and actual values for the natural gas price with XGBoost for the time period 1/1/2017 - 1/6/2020.
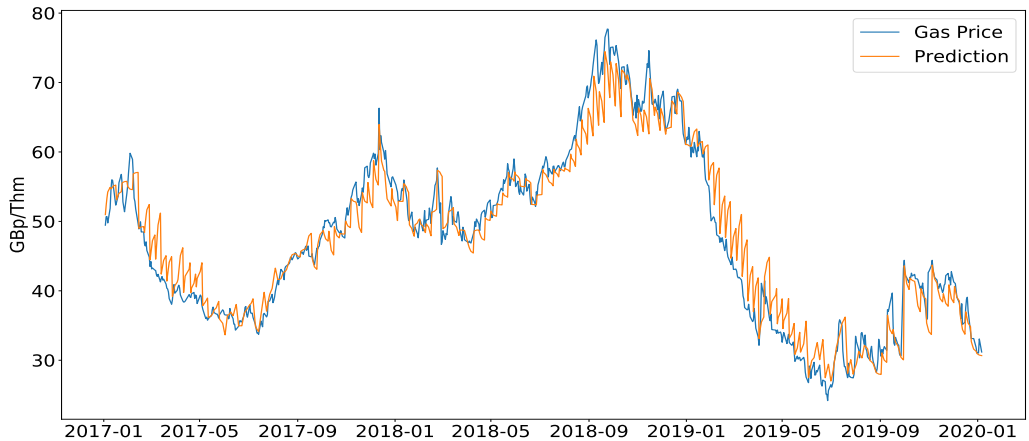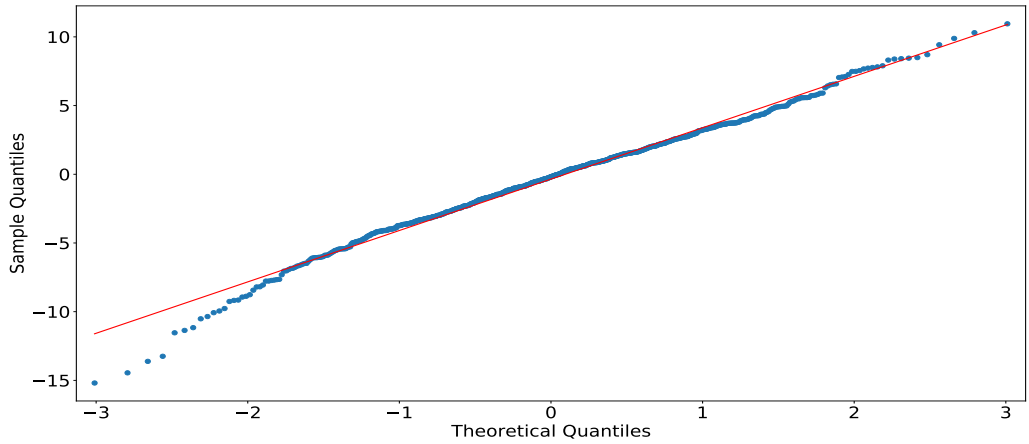

(c) The predicted and actual values for the natural gas price with LSTM for the time period 1/1/2017 - 1/6/2020.
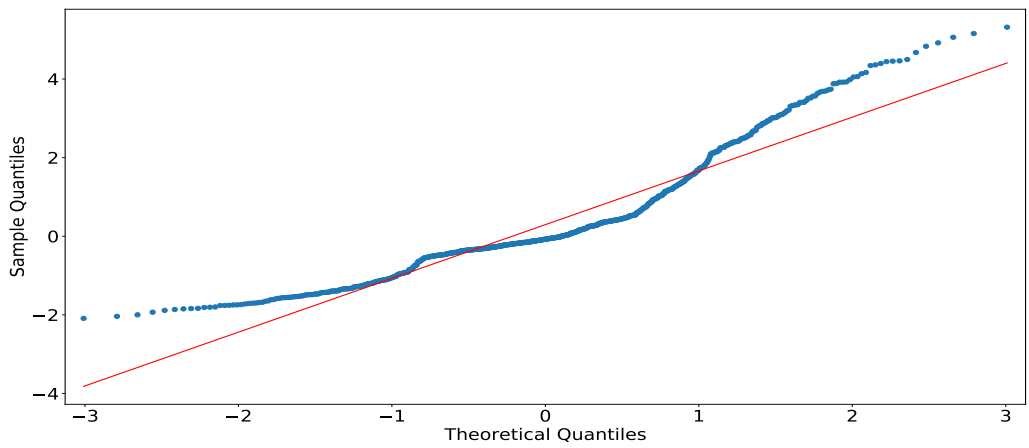
Figure 17: The 3 different models prediction of the natural gas price between 1/1/2017 - 1/6/2020.

(a) Q-Q plots for MLR

(b) Q-Q plots for XGBoost

(c) Q-Q plots for LSTM

Figure 18: The 3 different models prediction of the natural gas price between 1/1/2017 - 1/6/2020.

Table 7: The table includes the RMSE and MAE error measures for the different models, for the time period 1/1/2017 - 1/6/2020.

Model	RMSE	MAE
MLR	6.9	5.6
XGBoost	3.5	2.7
LSTM	3.1	2.1

The Q-Q plots in figure 18, shows that the errors from the MLR and the XGBoost was normally distributed, while the LSTM deviated from the line indicating non-normality in the distribution.
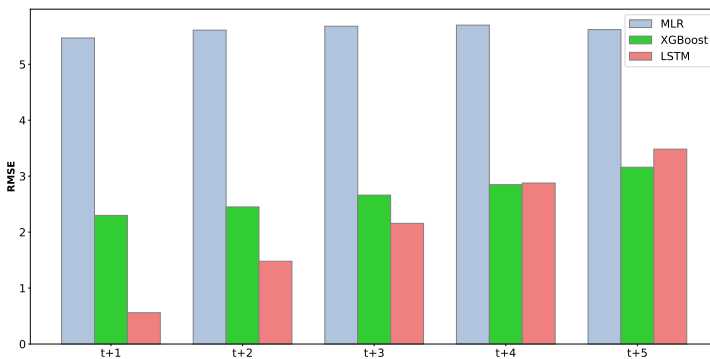

Figure 19: Mean Average Error for each time step the natural gas price prediction between 1/1/2017 - 1/6/2020.

When examining the prediction errors over successive time steps from the initiation of each forecast window in figure 23, it became evident that both XGBoost and MLR models demonstrated comparable levels of error for each individual step. This outcome was expected, considering how they are designed to build relationships between the gas price and relevant features. Conversely, the LSTM model demonstrated a progressive increase in error magnitude as the prediction moved further away from the most recent known data point. This outcome aligns with the anticipated behavior of all neural network models, as they are designed to retain and recognize patterns and structures. Consequently, as the temporal distance expanded, the model's ability to accurately forecast diminished, leading to a larger margin of error. It is worth noting that a slight elevation in error could also be observed for the XGBoost model, which could be attributed to the decreasing relevance of time-related features with each successive time step.

Table 8: The table includes the estimated regression coefficients for the final MLR model during the normal period. In addition, it displays the standard error, the t-value, p-value and the 95% confidence interval for the coefficients.

Variable	Coefficient	Std. Err.	t-value	P-value	[0.025, 0.975]
const	2.9644	0.908	3.266	0.001	1.185, 4.744
CoalPrice	0.5673	0.009	60.137	0.000	0.549, 0.586
CWV	-0.6673	0.037	-17.935	0.000	-0.740, -0.594

Table 8 shows that the coefficient for the coal price was positive and the CWV was negative. All variables were significant since the p-value was less than 0.05.
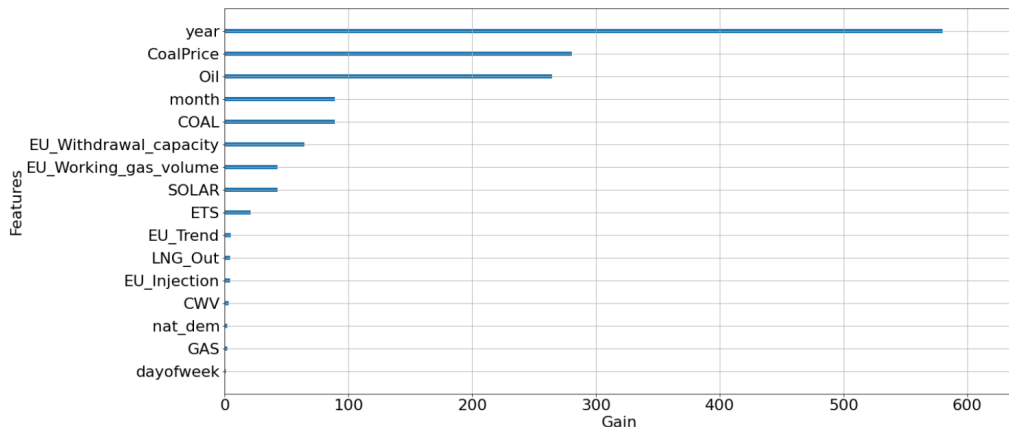

Figure 20: Gain for XGBoost during the normal period. Gain, as defined in equation 12, represents the aggregated accuracy improvement contributed by each feature.

The variable "year" exhibited the highest improvement in accuracy, which could be attributed to the temporal characteristics inherent in the gas market. The inclusion of this variable allowed the model to effectively capture and incorporate time-related patterns and trends, thereby enhancing its predictive performance. Additionally, the variables "Coal price" and "Oil price" demonstrated meaningful contributions to the model's accuracy as they pertained to other energy sources. Given their relevance to the gas market and their interconnectedness, these variables offered valuable information that aided in refining the model's predictive capabilities.

8.4 The 2021 natural gas supplier crisis and the Russian invasion

In August of 2021, the natural gas price in the UK started to drastically increase unlike anything seen before[37]. The principal contributor to this escalation was the worldwide upswing in demand as the world recuperated from the economic downturn resulting from the COVID-19 pandemic, with notable energy demand in Asia. Additionally, several other factors further increased the surge in gas prices. Among these were reduced gas supply from Russia to the European markets, an increase in gas demand for electricity generation during the summer of 2021 due to a sequence of nuclear power outages and the shutdown of the HVDC Cross-Channel interconnection following a fire.

Furthermore, the weather conditions worsened the situation. A cold winter during 2020/21 gave a greater need of natural gas for central heating than normal, resulting in further depletion of gas storages. The United Kingdom was further impacted by one of the least windy summers since 1961, resulting in severely reduced wind power generation[37].

The start of the Russian invasion of Ukraine took place in late February 2022, resulting in a record high price surge due to Russia being the largest exporter of natural gas to Europe [25].

From a modeling perspective, it is of great benefit to investigate and analyze these exceptional situations to uncover limitations and capabilities of the models.
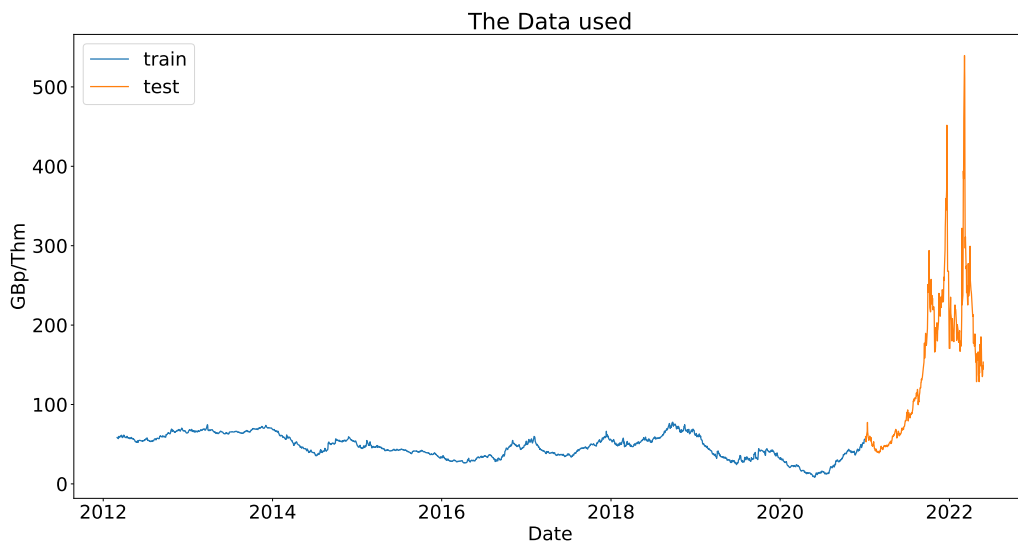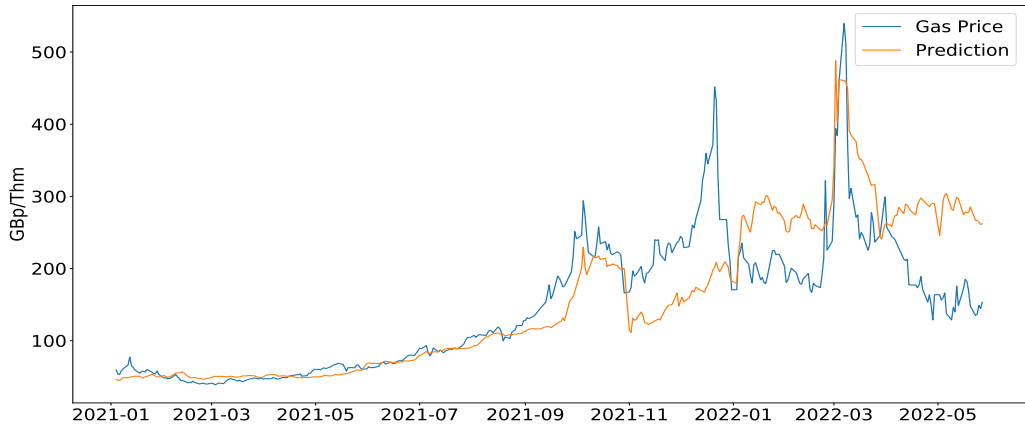

Figure 21: The training and test data for for the volatile period from 1/1/2009 - 30/5/2022.

(a) The predicted and actual values for the natural gas price with MLR for the time period 4/1/2021 - 30/5/2022.
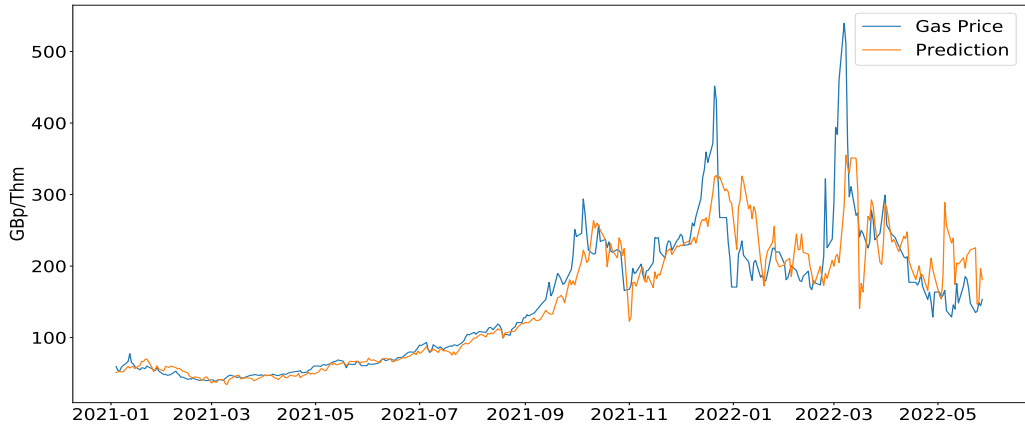
(b) The predicted and actual values for the natural gas price with XGBoost for the time period 4/1/2021 - 30/5/2022.
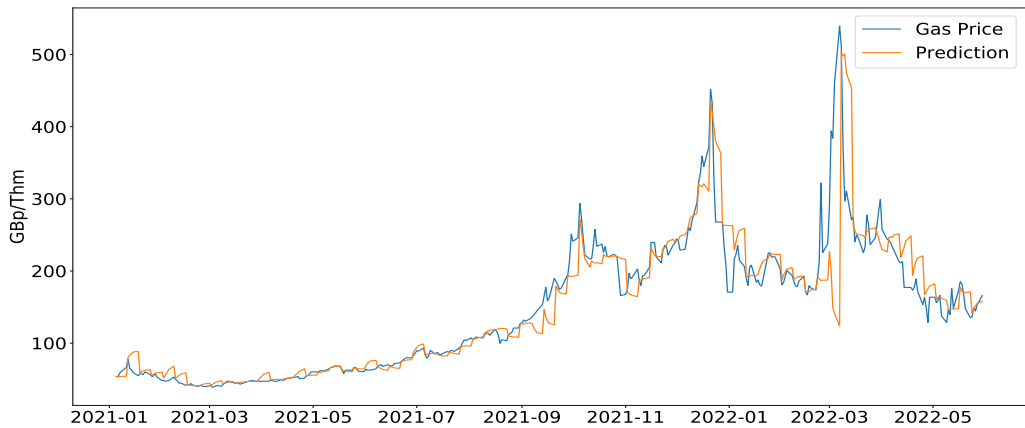
(c) The predicted and actual values for the natural gas price with LSTM for the time period 4/1/2021 - 30/5/2022.

Figure 22: The 3 different models prediction of the natural gas price between 4/1/2021 - 30/5/2022.

Table 9: The table includes the RMSE and MAE error measures for the different models, for the time period 4/1/2021 - 30/5/2022.

Model	RMSE	MAE
MLR	61.6	42.0
XGBoost	41.4	23.4
LSTM	40.6	20.4

The prediction results from the models in figure 22 and the errors in Table 9, exhibited interesting dynamics of the models. The MLR successfully detected the initial and third instances of price surges, although with inaccurate magnitudes. However, it struggled to capture the second surge accurately and displayed significant deviations both prior to and after. After the occurrence of the third surge, the model remained influenced by the preceding high surges. In contrast, the XGBoost algorithm successfully modeled the dynamics of all three surges, even though, like the MLR, it gave an inaccurate estimation of the magnitude of the last two surges. Finally, the LSTM method had the best fit, but missed the first part of the third price surge.
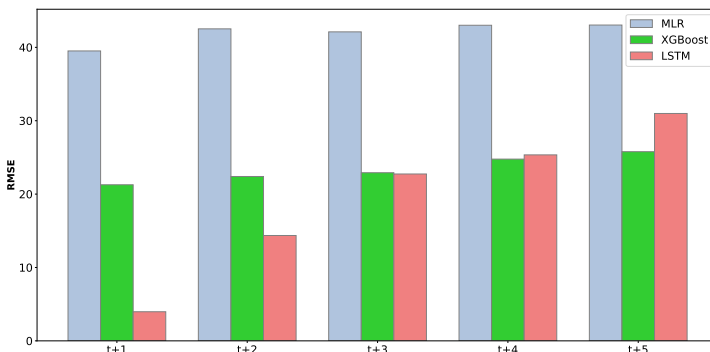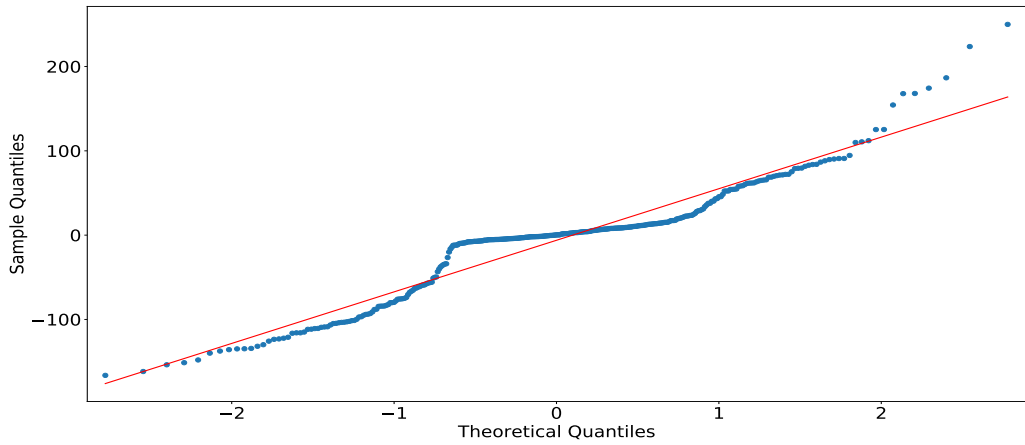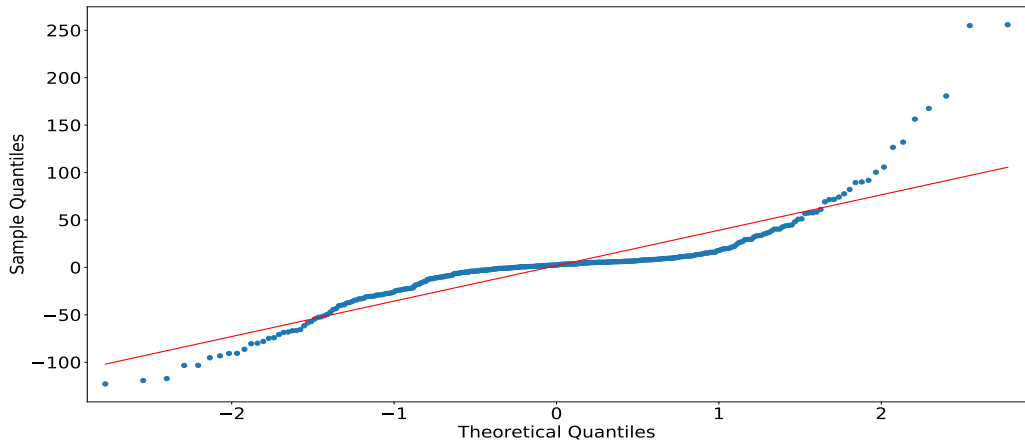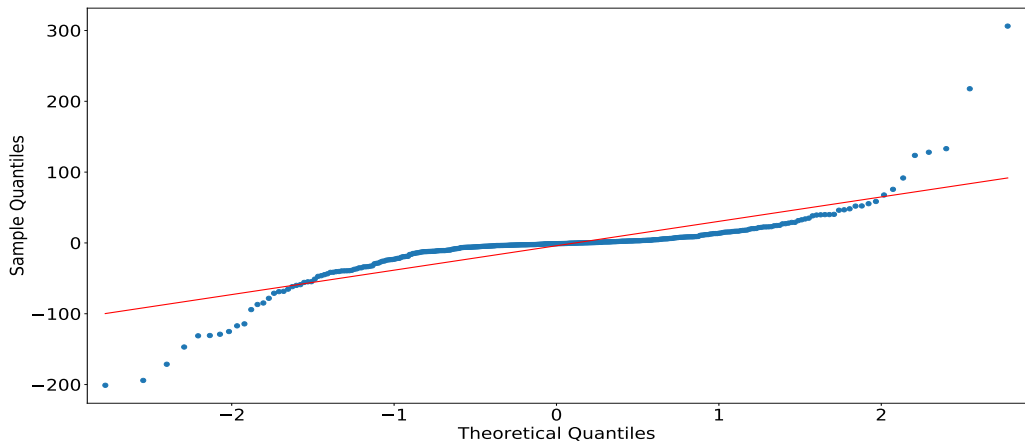

Figure 23: Mean Average Error for each time step the natural gas price prediction between 4/1/2021 - 30/5/2022.

(a) Q-Q plots for MLR

(b) Q-Q plots for XGBoost

(c) Q-Q plots for LSTM

Figure 24: The Q-Q plots for the 3 different models residuals prediction of the natural gas price between 4/1/2021 - 30/5/2022.

The Q-Q plots in figure 24 show that the residuals to all three models were not normally distributed.

Table 10: The table includes the estimated regression coefficients for the final MLR model during the volatile period. In addition, it displays the standard error, the t-value, p-value and the 95% confidence interval for the coefficients.

Variable	Coefficient	Std. Err.	t-value	P-value	[0.025, 0.975]
const	99.1434	4.262	23.260	0.000	90.785, 107.502
CoalPrice	0.6474	0.013	48.707	0.000	0.621, 0.673
CWV	-7.3947	0.332	-22.272	0.000	-8.046, -6.744

Table 10 shows the coefficient for the coal price being positive and the CWV being negative. The CWV value is significantly greater compared to the normal period. All variables are significant since the P-value < 0.05.
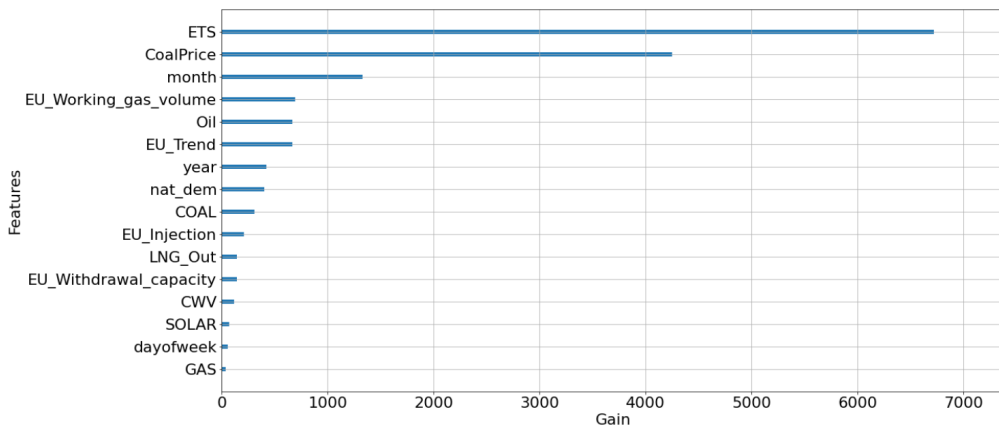
Figure 25: Gain for XGBoost during the volatile period. Gain, as defined in equation 12, represents the aggregated accuracy improvement contributed by each feature.

Figure 25 shows the gain for the XGBoost model, where the Emission Trading Scheme variable was significantly more important during the volatile period compared to the normal period, The coal price and the price of carbon in the ETS showed record high numbers, which was of no surprise as they were the most important covariates for the model.

8.5 Discussion

To facilitate the comparison of different modeling approaches, it is essential to ensure a fair starting point for each model. However, given the different nature of models and their approaches, achieving this fairness is often unfeasible. This is primarily due to differences in the models' objectives, which make comparisons unfair. For example, the MLR and XGBoost aims to construct a model that best fits the entire data set, while Long Short-Term Memory LSTM models focus on achieving the best possible prediction score. To reduce this difference between the models, specific approaches were employed to MLR and XGBoost. In the case of MLR, the observations were weighted, assigning greater importance to more recent data points. XGBoost utilized time-related features, such as year, month, day etc. to enable the model to model the temporal patterns and changes over time. Additionally, the LSTM also had an advantage in using the last observed natural gas price as a starting point to predict the consecutive 5 days, giving it an advantage over the MLR and XGBoost models.

Despite its limited predictive power, the MLR model offers valuable statistical inference, which can be of more value than predictive power in some cases. The model's simplicity also makes it understandable for non-statisticians. In contrast, the LSTM functions as a black box, where only a prediction is returned by feeding the algorithm with data, making it challenging to perform inference. The XGBoost model lies between these two models, with it being challenging to conceptually understand how changes in one variable would affect the prediction, even though different metrics on variable effectiveness, such as gain, can be obtained.

When selecting a modeling approach, the predictability of the model is a crucial factor to consider. In comparison to the LSTM and XGBoost models, the MLR model performed worse in some scenarios due to several reasons. For instance, the natural gas market is continuously evolving, and using a MLR model can be disadvantageous since the market ten years ago significantly differs from the present. Although the MLR model was trained using a weighted bias, the gas market is affected by different variables that undergoes changes at different rates and degrees, making it a suboptimal approach since it is difficult to individualize these weights. Another contributing factor is the limitation of variables the model can use. While XGBoost and LSTM can leverage all available variables without compromising its performance, the MLR model encounters issues related to multicollinearity, as many variables are correlated. This hinders the MLR ability to capture certain dynamics, which became evident when the model failed to capture the price dynamics around the beginning of 2022. During this period, increased demand for gas in Asia resulted in a surge in demand for global LNG. This information was not incorporated into any of the models. However, while XGBoost and LSTM can effectively utilize multiple variables, allowing marginal increases in several variables to compensate to some extent for the absence of information regarding the Asian market, the MLR model lacks this capability.

The XGBoost model performed significantly better than the MLR model regarding the RMSE and MAE and followed the price dynamic to a satisfactory extent. The LSTM model performed the best with the lowest RMSE and MAE, although not by much during the volatile period. Examining the plots, the fit was better for the LSTM. The reason for the small difference was that the LSTM failed to predict the beginning of the extreme price rise in march 2022, resulting in a large error for a few specific prediction points,

which contributed to an elevation in the overall average error of the model.

An advantage XGBoost and LSTM models have is that they are designed to filter out variables that do not improve predictions. However, if a variable's importance increases over time, both models can incorporate this change in their predictions. The name Long-Short Term Memory highlights this concept, where the short-term memory component possesses the capability to promptly assign greater significance to recently influential variables. The XGBoost model, on the other hand, constructs trees that adapt to changes over time with the given time features and therefore can accommodate such changes.

The MLR model outperformed the XGBoost and LSTM models during the price surges due to its ability to extrapolate. The XGBoost and LSTM models both struggled to predict the new highs of the gas price. The reason for their failure on these events was that they were not designed to be capable of extrapolating values. XGBoost, and all tree-based methods, are inherently incapable of effectively modeling a continuous function, consequently hindering their ability to generate predictions beyond the range of the training data. On the other hand, LSTM networks are primarily designed to capture and retain historical patterns, which naturally results in poor extrapolation capabilities.

The complexity of the model is also a crucial factor to consider. MLR models are relatively simple and easy to implement, making them a popular choice for many applications. However, their simplicity can limit their ability to capture complex relationships in the data. In contrast, XGBoost and LSTM models are designed to handle complex data and can capture non-linear relationships between variables. However, their complexity makes them more challenging to implement, also requiring significant computational resources and parameter tuning, which can lead to overfitting.

9 Conclusion and Future Work

In conclusion, this paper aimed to compare the three models: XGBoost, LSTM, and Multiple Linear Regression tasked with predicting the natural gas price in the UK. The evaluation was conducted over two distinct periods: a normal period spanning from 2017 to 2020, and a volatile period from 2021 to the middle of 2022.

During the normal period, both the XGBoost and LSTM models exhibited comparable levels of satisfactory prediction accuracy, surpassing the performance of the MLR model, which demonstrated adequate results. However, in the volatile period characterized by the extraordinary price surges, the MLR model displayed a superior ability to capture these price spikes due to its extrapolation capabilities. Conversely, the XGBoost model failed to accurately capture the magnitude of the price spikes, because it had not encountered such high gas prices during training. Similarly, the LSTM model encountered difficulties in accurately predicting the price spikes due to the same lack of prior exposure. However, the LSTM model showcased a faster adaptability to new patterns, enabling it to respond more swiftly to the price spikes which resulted in a comparatively lower overall score.

Furthermore, it is important to acknowledge that modeling can have objectives beyond the sole pursuit of prediction accuracy. In this regard, both the XGBoost and MLR models possess the capacity to quantitatively assess the relationships and significance of diverse features, thus facilitating a deeper understanding of the underlying dynamics. In contrast, the LSTM model functions as a "black box," solely providing predictions without additional interpretability or insights into the underlying mechanisms.

Going further, it would be interesting to explore alternative commodities to investigate potential variations in results. A significant challenge in modeling commodities lies in the multitude of variables influencing their prices, making it harder, if not impossible, to obtain a comprehensive set of variables. Alternatively, one could direct attention towards a different and more closed off gas market, such as the American natural gas market. Given its geographical isolation and self-sufficiency in gas production, the American market presents a reduced number of factors compared to the more expansive European or Asian gas markets. This narrower focus could provide valuable insights and potentially mitigate some of the complexities associated with modeling commodity prices. Further research could also explore hybrid approaches or alternative models that combine the strengths of these techniques to enhance prediction accuracy while accommodating some interpretability.

Bibliography

[1]   Jimmy Abualdenien and Andre Borrmann. 'Ensemble-learning approach for the classification of Levels Of Geometry (LOG) of building elements'. In: *Advanced Engineering Informatics* 51 (Jan. 2022), p. 101497. DOI: 10.1016/j.aei.2021.101497.

[2]   The U.S. Energy Information Administration. *Short-Term Energy Outlook Supplement: Market Drivers and Other Factors Affecting Natural Gas Prices.* URL: https://www.eia.gov/outlooks/steo/special/supplements/2023/2023_sp_01.pdf (visited on 5th June 2023).

[3]   U.S. Energy Information Administration. *Country Analysis Executive Summary: United Kingdom.* URL: https://www.eia.gov/international/content/analysis/countries_long/United_Kingdom/pdf/uk.pdf (visited on 5th June 2023).

[4]   Environmental protection agency. *The EU Emissions Trading System.* URL: https://www.epa.ie/our-services/licensing/climate-change/eu-emissions-trading-system-/ (visited on 10th Dec. 2022).

[5]   Michael Allen. *Understanding Regression Analysis.* Springer, 1997.

[6]   Øyvind Alsvik and Kari Barkved Strand. 'Sammenhengen mellom spot- og forward-priser i gassmarkedet - En empirisk studie av Title Transfer Facility og Zeebrugge'. In: *DET SAMFUNNSVITENSKAPELIGE FAKULTET, HANDELSHØGSKOLEN VED UIS MASTEROPPGAVE* (2014), p. 6.

[7]   Vikesh Singh Baghel. *Math behind GBM and XGBoost.* URL: https://medium.com/analytics-vidhya/math-behind-gbm-and-xgboost-d00e8536b7de (visited on 10th Dec. 2022).

[8]   C3.ai. *Gradient-Boosted Decision Trees (GBDT).* URL: https://c3.ai/glossary/data-science/gradient-boosted-decision-trees-gbdt/ (visited on 8th June 2023).

[9]   INCH CALCULATOR. *Million BTU to Therms Converter.* URL: https://www.inchcalculator.com/convert/million-btu-to-therm/ (visited on 12th June 2023).

[10]   Data Science Central. *5 Most Popular Measures For Capturing Forecasting Error For Time Series.* URL: https://www.datasciencecentral.com/5-most-popular-measures-for-capturing-forecasting-error-for-time/ (visited on 10th June 2023).

[11]   McKinsey Company. *Can carbon prices fire up gas demand in electricity generation?* URL: https://www.mckinsey.com/industries/oil-and-gas/our-insights/can-carbon-prices-fire-up-gas-demand-in-electricity-generation (visited on 8th June 2023).

[12]   DanB. *XGBoost.* URL: https://www.kaggle.com/code/dansbecker/xgboost (visited on 15th Dec. 2022).

[13]   Jalal Dehnavi, Franz Wirl and Yuri Yegorov. 'Arbitrage in natural gas markets?' In: *International Journal of Energy and Statistics* 3 (2015). DOI: 10.1142/S2335680415500180.

[14]   Meteostat Developers. *Meteostat.* URL: https://dev.meteostat.net/python/#installation.

[15]   EEA. *EU Emissions Trading System (ETS) data viewer.* URL: https://www.eea.europa.eu/data-and-maps/dashboards/emissions-trading-viewer-1.

[16]   Joint Office of Gas Transporters. *Composite Weather Variable: Temperature and Weather Analysis.* URL: https://www.gasgovernance.co.uk/sites/default/files/ggf/2019-03/CWV%5C%20-%5C%20Temp%5C%20and%5C%20Wind%5C%20Speed%5C%20Analysis.pdf (visited on 8th June 2023).

[17] GIE. *AGSI*. URL: https://agsi.gie.eu/.

[18] GIE. *ALSI*. URL: https://agsi.gie.eu/.

[19] CME Group. *Are Crude Oil Natural Gas Prices Linked?* URL: https://www.cmegroup.com/education/articles-and-reports/are-crude-oil-natural-gas-prices-linked.html (visited on 8th June 2023).

[20] heatable. *TWhere Does the UK Get its Gas From? Official Statistics Explored*. URL: https://heatable.co.uk/boiler-advice/where-does-the-UK-get-its-gas-from (visited on 5th June 2023).

[21] Ahmed Hendy. *How fast does natural gas travel through pipelines?* URL: https://www.quora.com/How-fast-does-natural-gas-travel-through-pipelines (visited on 10th Dec. 2022).

[22] IBM. *What is supervised learning?* URL: https://www.ibm.com/topics/supervised-learning (visited on 12th June 2023).

[23] Investing.com. *Commodities*. URL: https://www.investing.com/.

[24] Investopedia. *Market Efficiency Explained: Differing Opinions and Examples*. URL: https://www.investopedia.com/terms/m/marketefficiency.asp (visited on 12th June 2023).

[25] J.P.Morgan. *How are sanctions against Russia impacting oil and gas prices?* URL: https://www.jpmorgan.com/insights/research/oil-gas-energy-prices (visited on 9th June 2023).

[26] Dive Into Deep Learning. *10.1. Long Short-Term Memory (LSTM)*. URL: https://d2l.ai/chapter_recurrent-modern/lstm.html (visited on 9th June 2023).

[27] Eric Luellen. *Mastering XGBoost*. URL: https://towardsdatascience.com/mastering-xgboost-2eb6bce6bc76 (visited on 12th June 2023).

[28] Amol Mavuduru. *Why XGBoost can't solve all your problems*. URL: https://towardsdatascience.com/why-xgboost-cant-solve-all-your-problems-b5003a62d12a (visited on 10th Dec. 2022).

[29] Vishal Morde. *XGBoost Algorithm: Long May She Reign!* URL: https://towardsdatascience.com/https-medium-com-vishalmorde-xgboost-algorithm-long-she-may-rein-edd9f99be63d (visited on 18th Dec. 2022).

[30] nationalgrid. *Gas Demand Forecasting Methodology*. URL: https://www.nationalgas.com/document/132516/download.

[31] nationalgrid. *Welcome to the National Grid ESO Data Portal*. URL: https://data.nationalgrideso.com/.

[32] NVIDIA. *XGBoost*. URL: https://www.nvidia.com/en-us/glossary/data-science/xgboost/ (visited on 18th Dec. 2022).

[33] OPTUNA. *optuna.samplers.GridSampler*. URL: https://optuna.readthedocs.io/en/v2.0.0/reference/generated/optuna.samplers.GridSampler.html (visited on 5th June 2023).

[34] Manimaran P. et al. 'Modelling Financial Time Series'. In: 2006, pp. 183–191. ISBN: 978-88-470-0501-3. DOI: 10.1007/978-88-470-0502-0_19.

[35] Power. *The Economics of Coal-to-Gas Switching*. URL: https://www.powermag.com/the-economics-of-coal-to-gas-switching/ (visited on 5th June 2023).

[36] Hantao Wang. *Identifying the Correlation between Temperature and Gas Consumption in a Local Energy System.* URL: http://www.csee.org.cn/u/cms/English/201809/25163124c5oc.pdf (visited on 5th June 2023).

[37] Wikipedia. *2021 United Kingdom natural gas supplier crisis.* URL: https://en.wikipedia.org/wiki/2021_United_Kingdom_natural_gas_supplier_crisis (visited on 9th June 2023).

[38] Wikipedia. *Artificial neural network.* URL: https://en.wikipedia.org/wiki/Artificial_neural_network (visited on 15th Dec. 2022).

[39] Wikipedia. *British thermal unit.* URL: https://en.wikipedia.org/wiki/British_thermal_unit.

[40] Wikipedia. *Gradient boosting.* URL: https://en.wikipedia.org/wiki/Gradient_boosting (visited on 8th June 2023).

[41] Wikipedia. *Liquefied natural gas.* URL: https://en.wikipedia.org/wiki/Liquefied_natural_gas.

[42] Wikipedia. *Liquefied natural gas.* URL: https://en.wikipedia.org/wiki/Liquefied_natural_gas (visited on 10th Dec. 2022).

[43] Wikipedia. *Long-Short Term Memory.* URL: https://en.wikipedia.org/wiki/Long-short-term_memory (visited on 9th June 2023).

[44] Wikipedia. *Natural gas storage.* URL: https://en.wikipedia.org/wiki/Natural_gas_storage (visited on 5th June 2023).

[45] Wikipedia. *Q–Q plot.* URL: https://en.wikipedia.org/wiki/Q%5C%E2%5C%80%5C%93Q_plot (visited on 11th Dec. 2022).

[46] Wikipedia. *Understanding LSTM Networks.* URL: https://colah.github.io/posts/2015-08-Understanding-LSTMs/ (visited on 9th June 2023).

[47] Wikipedia. *Variance inflation factor.* URL: https://en.wikipedia.org/wiki/Variance-inflation_factor (visited on 15th Dec. 2022).

[48] Charles Zaiontz. *Method of Least Squares for Multiple Regression Detailed.* URL: https://www.real-statistics.com/multiple-regression/least-squares-method-multiple-regression/least-squares-method-multiple-regression-detailed/ (visited on 10th Dec. 2022).

Appendix

The derivatives for the parameters in the LSTM model are calculated by first defining the loss function

$$J_t = \frac{1}{2}(\hat{y}_t - y_t)^2.$$

The derivatives are calculated following the equations

$$\frac{\partial J_t}{\partial H_t} = \frac{\partial J_t}{\partial \hat{y}_t} = \hat{y}_t - y_t = J_\Delta$$

$$\frac{\partial J}{\partial O_t} = \frac{\partial J}{\partial H_t} \cdot \frac{\partial H_t}{\partial O_t} = J_\Delta \cdot \tanh(C_t)$$

$$\frac{\partial J}{\partial F_t} = \frac{\partial J}{\partial H_t} \cdot \frac{\partial H_t}{\partial F_t} = J_\Delta \cdot O_t \cdot (1 - \tanh(C_t)^2) \cdot C_{t-1}$$

$$\frac{\partial J}{\partial C_t} = \frac{\partial J}{\partial H_t} \cdot \frac{\partial H_t}{\partial C_t} = J_\Delta \cdot O_t \cdot (1 - \tanh(C_t)^2)$$

$$\frac{\partial J}{\partial I_t} = \frac{\partial J}{\partial H_t} \cdot \frac{\partial H_t}{\partial I_t} = J_\Delta \cdot O_t \cdot (1 - \tanh(C_t)^2) \cdot \widetilde{C}_t$$

$$\frac{\partial J}{\partial \widetilde{C}_t} = \frac{\partial J}{\partial H_t} \cdot \frac{\partial H_t}{\partial \widetilde{C}_t} = J_\Delta \cdot O_t \cdot (1 - \tanh(C_t)^2) \cdot I_t$$

$$\frac{\partial J}{\partial C_{t-1}} = \frac{\partial J}{\partial H_t} \cdot \frac{\partial H_t}{\partial C_{t-1}} = J_\Delta \cdot O_t \cdot (1 - \tanh(C_t)^2) \cdot F_t.$$

These are used to obtain the derivatives for the weights

$$\frac{\partial J}{\partial W_O} = \frac{\partial J}{\partial O_t} \cdot \frac{\partial O_t}{\partial W_O} = J_\Delta \cdot \tanh(C_t) \cdot \sigma(a_O)(1 - \sigma(a_O)) \cdot Z_t$$

$$\frac{\partial J}{\partial W_F} = \frac{\partial J}{\partial F_t} \cdot \frac{\partial F_t}{\partial W_F} = J_\Delta \cdot O_t \cdot (1 - \tanh(C_t)^2) \cdot C_{t-1} \cdot \sigma(a_F)(1 - \sigma(a_F)) \cdot Z_t$$

$$\frac{\partial J}{\partial W_I} = \frac{\partial J}{\partial I_t} \cdot \frac{\partial I_t}{\partial W_I} = J_\Delta \cdot O_t \cdot (1 - \tanh(C_t)^2) \cdot \widetilde{C}_t \cdot \sigma(a_I)(1 - \sigma(a_I)) \cdot Z_t$$

$$\frac{\partial J}{\partial W_{\widetilde{C}_t}} = \frac{\partial J}{\partial \widetilde{C}_t} \cdot \frac{\partial \widetilde{C}_t}{\partial W_{\widetilde{C}_t}} = J_\Delta \cdot O_t \cdot (1 - \tanh(C_t)^2) \cdot I_t \cdot \sigma(a_I)(1 - \sigma(a_I)) \cdot Z_t$$

and for the biases

$$\frac{\partial J}{\partial b_O} = \frac{\partial J}{\partial O_t} \cdot \frac{\partial O_t}{\partial b_O} = J_\Delta \cdot \tanh(C_t) \cdot \sigma(a_O)(1 - \sigma(a_O))$$

$$\frac{\partial J}{\partial b_F} = \frac{\partial J}{\partial F_t} \cdot \frac{\partial F_t}{\partial b_F} = J_\Delta \cdot O_t \cdot (1 - \tanh(C_t)^2) \cdot C_{t-1} \cdot \sigma(a_F)(1 - \sigma(a_F))$$

$$\frac{\partial J}{\partial b_I} = \frac{\partial J}{\partial I_t} \cdot \frac{\partial I_t}{\partial b_I} = J_\Delta \cdot O_t \cdot (1 - \tanh(C_t)^2) \cdot \widetilde{C}_t \cdot \sigma(a_I)(1 - \sigma(a_I))$$

$$\frac{\partial J}{\partial b_{\widetilde{C}_t}} = \frac{\partial J}{\partial \widetilde{C}_t} \cdot \frac{\partial \widetilde{C}_t}{\partial b_{\widetilde{C}_t}} = J_\Delta \cdot O_t \cdot (1 - \tanh(C_t)^2) \cdot I_t \cdot \sigma(a_I)(1 - \sigma(a_I)).$$