Thomas Hjertner
Johan Gundersen Høyslo

Analysis of the Norwegian power market with prediction of day-ahead spot prices by means of Linear Regression and Machine Learning.

Bachelor's thesis in Business Administration, Business Analytics
Supervisor: Denis Becker
April 2022

Bachelor's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Economics and Management
NTNU Business School

NTNU
Norwegian University of
Science and Technology

Thomas Hjertner
Johan Gundersen Høyslo

Analysis of the Norwegian power market with prediction of day-ahead spot prices by means of Linear Regression and Machine Learning.

Bachelor's thesis in Business Administration, Business Analytics
Supervisor: Denis Becker
April 2022

Norwegian University of Science and Technology
Faculty of Economics and Management
NTNU Business School

NTNU
Norwegian University of Science and Technology

PREFACE

Over two and a half years ago, we started this education with little to no idea of how our interest would shift from traditional finance and business toward more modern takes on the traditional subject of business. Moreover, last year we explored the subject of information technology and suddenly noticed a deep interest in the combination of business and programming. Now, we have realized that business, economics, and finance are heading into the digital world of implementation of programming and machine learning.

We live in a world in constant motion, whether regarding financial markets or geopolitical events. Over the last three years, we have seen natural disasters, a global pandemic, and now war. As students at this point in time, it is our responsibility to gain the skills required to navigate through this fast-moving world. With the knowledge of how to combine business, economics, and finance with computer science, we feel better prepared for the rapid pace of society, and the changing world of business.

Financial markets are also due to a digital transformation. Stock markets globally are embossed by advanced trading algorithms built on statistical foundations, and machine learning is utilized to benefit financially from the forecasts that these models yield. These techniques correspondingly emboss the electricity market.

We have found working on this bachelor thesis very intriguing. Both authors are confident that the major of Business Analytics is one we also want to pursue at masters level next year. We want to express gratitude to our supervisor Denis Becker for giving us the foundation needed to write this thesis and guidance when writing the thesis. The course Essentials of Business Analytics (BBAN3001) is a definite highlight of our three-year study and is a course that sparked our interest in the field. We also want to thank NordPool for providing us with free access to their FTP server used to get data for the thesis.

ABSTRACT

Electricity differs from most other tradeable commodities in that it cannot easily be stored. Combining econometric and machine learning methods in the electricity market can help predict where there is a balance between generation and consumption, as well as what the price will be. The dominant methods used to provide forecasts are severely advanced machine learning models. As different methods are built on different assumptions, one must be able to navigate and adapt to the type of data available. We have used data from various open- and private sources to utilize both statistical and machine learning approaches to a) analyze and b) forecast the electricity market. The result of our study yields a some-what precise machine learning model, that to some degree can forecast the electricity price.

Our findings have resulted in a deeper understanding of what drives the electricity price, which we find essential to make forecasts. The variables are chosen manually, and looking backwards, might have led to some reliability issues as it may have introduced selection bias to the dataset.

SAMMENDRAG

Elektrisk strøm skiller seg fra de fleste andre omsettelige råvarer ved at den egner seg dårlig til lagring. Ved å kombinere økonometri og maskinlæringsmetoder i kraftmarkedet kan man forsøke å predikere hvor det er balanse mellom produksjon og forbruk, samt forsøke å predikere prisen. De dominerende metodene som brukes for å gi prognoser er svært avanserte maskinlæringsmodeller. Ettersom de ulike metodene er bygget på ulike forutsetninger, må man kunne navigere og tilpasse seg typen data som er tilgjengelig. Vi har brukt data fra ulike åpne og private kilder for å benytte både statistiske og maskinlæringstilnærminger for å a) analysere og b) predikere kraftmarkedet. Resultatet av oppgaven er en noe presis maskinlæringsmodell, som til en viss grad er i stand til å forutsi strømprisen.

Funnene våre har resultert i en dypere forståelse av hva som driver strømprisen, noe vi finner avgjørende for å lage prognoser. Variablene som blir benyttet ble valgt manuelt, noe som sett i ettertid kan ha ført til reliabilitetsproblemer på grunn av utvalgsskjevhet.

LIST OF FIGURES AND TABLES

TABLE OF CONTENT

1. INTRODUCTION

Electricity is, in economic terms, defined as a tradeable commodity, and power exchanges enable electricity to be traded through bids and buys. In practical applications, electricity describes the measurement of energy flow through a metered point in a given period and is measured in megawatt-hours (MWh). On the other hand, power is the metered net electrical rate of transfer at any given point in time, thus measured in megawatts (MW) (U.S. Energy Information Administration, 2021).

Electricity is a relatively new type of tradeable commodity. Several characteristics differentiate electricity from commodities like crude oil or natural gas. Firstly, in comparison, electricity is completely fungible; thus, one MWh of electricity produced from either natural gas or coal contains the same amount of energy (CME Group, 2017). Secondly, it must be produced and consumed simultaneously as industrial size battery storage lacks efficiency and is prohibitable expensive. Thus, supply must meet demand precisely in the power grid. Ancillary services and demand response programs, such as *The Energy Act*, ensures that an increase in demand is met in real-time. This supports that uneconomic generation resources are not dispatched when demand is low, as they are only put into service during peak demand (CME Group, 2017).

Norway is part of an integrated, Nordic power market. The interconnection of the power market has contributed to strengthening the supply capacity, reducing overall costs in the power supply, and facilitating more renewable power production in the Nordic region of Europe (NordReg, 2019). The Nordic power market is joined by Norway, Sweden, Denmark, and Finland and is, in its turn, additionally integrated with the European market through interconnectors in the Netherlands, Germany, Poland, the Baltics, and pre-war Russia. In 2021 two additional connectors were established and preceded to start operations from Norway to intercontinental Europe through Germany and the UK, respectively, the Nord Link- and North Sea Link cable.

Most Nordic electricity is generated from hydro-, nuclear- and wind-powered turbines, whereas hydro-generated electricity is primarily produced in Norway (NordReg, 2019). This means that most of the generation in the Nordic countries comes from renewable resources and makes the Nordic power system consists of a mixture of generation sources reflected by the country's energy-intensive industry. Compared to global consumption and total power usage, Norway and the Nordics consume notably more electricity than the global average (Ministry of Petroleum and Energy, 2021).

1.1 ACTUALIZATION

Throughout 2021, most citizens in Norway had their eyes shut wide open as the price of electricity suddenly increased to an average of NOK 0,509/kWh in Q2 2021, not taking fees and grid rent into consideration (Tørmoen, 2022). Compared to the same period last year, the price increased by more than 300%. The electricity prices in 2021 reached all-time highs compared to the last decades, impacting Norwegian families as well as small- and bigger-size businesses. Neither does it seems like the prices are about to decrease and normalize to lower levels in the coming periods. In an interview earlier this year, the chief executive officer of the Norwegian Water Resources and Energy Directorate (NVE) said that the electricity prices are assumed to stay at these high levels at least for the whole of 2022 (Fosse, 2022). Furthermore, in the same interview, Fosse predicted that the prices we will see this summer will be likewise those we usually observe during winters.

The electricity market is volatile and non-stationary. Thus, the electricity prices and consumption are in constant motion and are expected to change over time. No reliable methods are discovered and shown to predict the fluctuation in the price, and forecasts and estimates can be resolved through both statistical and machine learning approaches.

1.2 THESIS STATEMENT

In this thesis, we want to analyze the electricity market thoroughly. We hope to find the reasoning behind the relatively sudden increase in the electricity price and discover the factors driving this price. *Using linear regression and machine learning, we will establish a model that can catch these factors and predict the day-ahead spot price*.

As the power market is primarily affected by endogenous and exogenous factors, it may be hard to distinguish between what may be regulated and within our control and the factors that cannot. Throughout the thesis, we will explore different perspectives through different analyses to understand this market. What seems to be the leading drivers of the electricity price? Does the increasing number of electricity lines to continental Europe affect the prices? These questions might be genuinely relevant these days and going forward. Trends in financial markets and deregulation may also be relevant factors to discuss. Therefore, we will consider history relevant to some extent but avoid the trap of inductive thinking.

The motivation behind our topic of choice, the electricity market, is its present relevancy. We also find the opportunity to combine economic studies with machine learning techniques highly relevant and beneficial. By analyzing our dataset containing daily data from 01.01.2009 until 31.12.2021, gathered from different viable sources, we can combine these two fields to analyze the electricity market efficiently and thoroughly. To summarize the key takeaways from the thesis statement:

What affects and drives the price of electricity?

We will analyze the electricity market utilizing linear regression and machine learning.

1.3 STRUCTURE

Chapter 1: **Introduction.**

Chapter one contains a brief introduction to the topic and a presentation of the thesis statement and object. Motivation and reasoning behind the choice of topic and methods are also presented in this section.

Chapter 2: **Literature/Theory.**

Chapter two contains the theoretical content we have based our methods on. It starts by presenting some relevant background information, then proceeds to introduce necessary information about the Nordic power market and exchange. After that, the theory behind the statistical- and machine learning approaches is presented. Lastly, the chapter ends with a brief introduction to the dataset, described in detail in appendix 7-1.

Chapter 3: **Methods.**

Chapter three will contain the methods used to gather the data, how the data is analyzed, and a further explanation of why specific methods and models used in the analysis were chosen. Limitations will also be mentioned. Lastly, the chapter ends with explaining the mathematics behind the models used in the thesis.

Chapter 4: **Results/discussion**

This chapter outlines the findings of our analysis and models. We will briefly explain the facts and results from our models and present illustrations to visualize these results using graphs and tables. After the results are presented, interpretation of the results will be discussed concerning the objective of the thesis.

Chapter 5: **Conclusion.**

In the fifth and final chapter we will highlight what we have achieved. The limitations mentioned in the methods chapter will also be considered to better ground our thesis. Our experiences and suggestion for further research will end this bachelor thesis.

2. THEORY AND LITERATURE

2.1 BACKGROUND

Over the last decades, the energy markets have shifted towards deregulation and harmonization. The harmonization resulted in the joint Nordic market of today. Deregulation led to establishing the framework for competition in both generation and trading of electricity by decentralized private organizations. In 1991 the legislation New Energy Act was effectuated. It provided a legal basis for regulating the grid companies to ensure that electricity was generated, converted, transmitted, traded, distributed, and used rationally and in the best interest of society (Grantham Research Institute on Climate Change and the Enviornment, 2018). It also made it possible to govern and regulate energy trading, cross-border interconnectors, district heating facilities, system operations, electricity supply, and contingency planning of power supply (Ministry of Petroleum and Energy, 2019).

The act only intended to regulate the increased competition; hence, the businesses used suitable decision support models to increase margins and reduce risk (Becker & Li, 2021). Accurate pricing forecast models became the industry standard to keep up with the competition and stay competitive. Because of market coupling, the features affecting the electricity price increased; hence the models to forecast the price became more complex.

2.2 NordPool

The Nordic electricity market described briefly in the introduction. is connected to the Nordic power exchange. The Nordic power exchange is among the most prominent European exchanges measured in transmission capacity. The marketplace is split into several sub-grids based on transmission lines and grid quality. The Norwegian grid is divided into five pricing areas, as seen in figure 2.1, and the rest of the Nordic market. Nordpool provides the coupling and cross-border trading (NordPool Group, 2022). Each country has its Transmission System Operator (TSO), which fills the role of forecasting and scheduling generation to ensure that sufficient generation and backup power is stored to meet unexpected demand or generation loss (CME Group, 2017). TSOs are important in the Nordic grid as the generation comes mostly from renewable resources, viable for production loss due to variations in weather.

In 2020, the annual NordPool trading volume was 995 TWh, whereas 717,9 TWh came from the Nordic and Baltic markets (NordPool Group, 2021) and equaled about 72% of the total trading volume and a value of approximately 10,895 billion Euros.

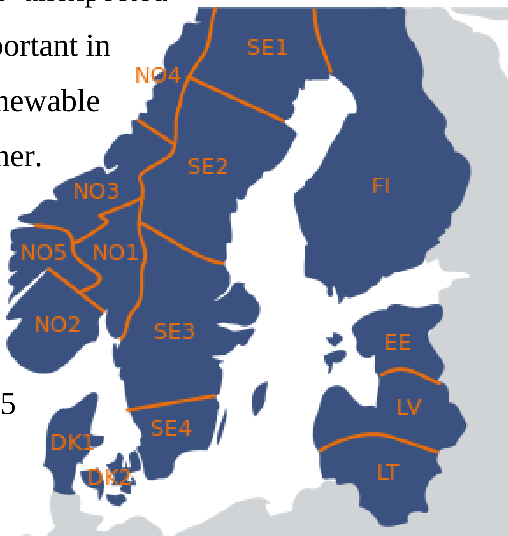

FIGURE 2-1: OVERVIEW OF THE NORDIC POWER MARKET

2.3 Trading at NordPool

NordPool delivers day-ahead, and intra-day trading, clearing, and settlement to customers. With intra-day trading means that customers who want to purchase or sell electricity must place their orders by midnight the day before the energy is supposed to be delivered to the grid.

The NordPool marketplace operates as auction-based on day-ahead trading where the electricity spot price is the element of interest. Clients who purchase or sell electricity must notify NordPool by midnight the day before the electricity is transmitted to the grid. Bids are digitally transmitted to the marketplace, generating a bid/demand curve based on purchase bids and sales offers. Hourly-, block- and flexible hour bids are the three types of bid methods utilized at NordPool. Participants must place their bids in the designated grid area (figure 2.1) where the electricity is produced or consumed. If a market participant operates a hydro plant in the NO1 zone, the operator must also sell the electricity in this zone.

Every day at 2.00 p.m., the price is announced for the next 12 to 36 hours. Each designated zone/area is given its calculated price before the system price is shared. The system price is a hypothetical shared Nordic price, and all the zones would have a shared system price given no transmission limitations. As a result of grid congestion and bottlenecks, prices in various zones frequently differ. From midnight to 11.00 p.m. the next day, there exists a 24-hour period wherein the end of the period, the market participants are informed of how much electricity they have either sold or bought after the pricing has been computed.

2.4 PREDICTING FUTURE SPOT PRICES

In recent years, the utilization of statistical methods using machine learning to predict the future value of variables has seen a rise in popularity. Access to data is a significant factor. When provided with large-scale data, particular time-series analyses can be conducted utilizing various methods and algorithms. Generally, electricity price forecasting can be divided into five categories: multi-agent, fundamental, reduced-form, statistical, and computational intelligence (CI) models. The latter, and more particular deep neural networks (DNNs), are considered the state-of-the-art method. DNNs can furthermore be categorized as forwarding neural networks (FNNs), recurrent neural networks (RNNs), and convolutional neural networks (CNNs), where each suits different means. The RNNs are most applicable to time-series predictions (Li & Becker, 2021). Furthermore, extended short-term memory networks are variants of RNNs capable of handling the vanishing gradient problem (Calin, 2020).

The statistical approach can predict the most relevant dynamics of the electricity price and map and generalize specific market mechanisms. Additionally, the regression model explores significant relationships between dependent variables and independent variables, indicating the strength of the independent variables' impact on the dependent variable, and makes it possible to compare the effect of variables measured on different scales (Panchotia, 2020). The regression line can predict the value of the price variable based on given predictor variables.

2.4.1 STATISTICAL APPROACH: MULTIVARIATE LINEAR REGRESSION WITH OLS

Assuming that the relationship between the data variables is linear, one can use linear regression to make assumptions about how and what variables affect the price of electricity. The linear relationships are estimated using Ordinary Least Squares (OLS). The fitted line in multiple linear regression is where the sum of squares between the observed value in the dataset and those predicted by the linear function is minimized. The estimates return the expected marginal effects of the variables, partialized from the effects by the others. Following is equations and variables making up the multivariate linear regression model:

$$y = \beta_0 + \beta_1 + x_1 + \cdots + \beta_k + x_k + u$$

y – dependent variable

x_k – independent variable (k)

β_0 – constant intercept variable

β_k – slope of the parameter x_k

u – noise

β_0 interprets the point on the y axis where the fitted line of all the x_k intercept ($x_k = 0$), β_0 is the constant intercept variable, while β_k is the slope of the parameter x_k. Lastly, the noise captures the rest of the variation in the dependent variable, excluded unabsorbed independent variables.

Robust standard errors are used in the MLR to obtain unbiased standard errors of OLS coefficients in case of heteroscedasticity. See 3.3.2 for more about the assumptions for MLR.

The standardized beta coefficient compares the strength of the effect of each independent variable (x) to the dependent variable (y). The higher the absolute value of the beta coefficient, the stronger the effect. For example, a beta of -0.9 has a more substantial effect than +0.8. Standardized beta coefficients have standard deviations as their units. The formula is:

$$\hat{\beta}_{y,k} = \frac{s_x}{s_y} \cdot \hat{\beta}_k$$

, where the s interprets the standard deviation of the depended and independent variables.

The standardized beta accumulates a number between -1and 1 and can therefore be used to compare variables with different levels of measurement.

T-values and p-values are used to test the significance of the observations in a multivariate linear regression model. These tests are translated into:

H_0: x has no effect on y (coefficient = 0)
H_1: x affects y (coefficient ≠ 0)

The formula of the t-value is:

$$t = \frac{\beta}{SE}$$

, and is tested against a critical value that accounts for the chosen significance level (α) and degrees of freedom (*df*). When df is higher than 120 and the level of significance reach 5%, the critical value is approximately 1,96.

P-values measure the possibility of not observing significant results. If $p < \alpha$, we reject the null hypothesis (H_0) and conclude with the alternative hypothesis (H_1). If $p < \alpha$, we fail to reject the H_0.

The models' output also shows the F-test score that indicates whether some (at least one) of the variables significantly differ from zero.

2.4.2 *ML APPROACH: LSTM – LONG SHORT-TERM MEMORY NETWORKS*

LSTM networks are categorized as an artificial recurrent neural network that is capable of learning long-term dependencies between time steps of sequence data (MathWorks, 2022). The LSTM-model is applicable for a variety of objects like classifying, processing and making predictions based on time series data. Contrary to standard feedforward neural networks, LSTM uses feedback connections such that make it capable of processing multiple datapoints and entire sequences of data. The LSTM network has built-in mechanisms that control how information is memorized or abandoned throughout time (Becker & Li, 2021).

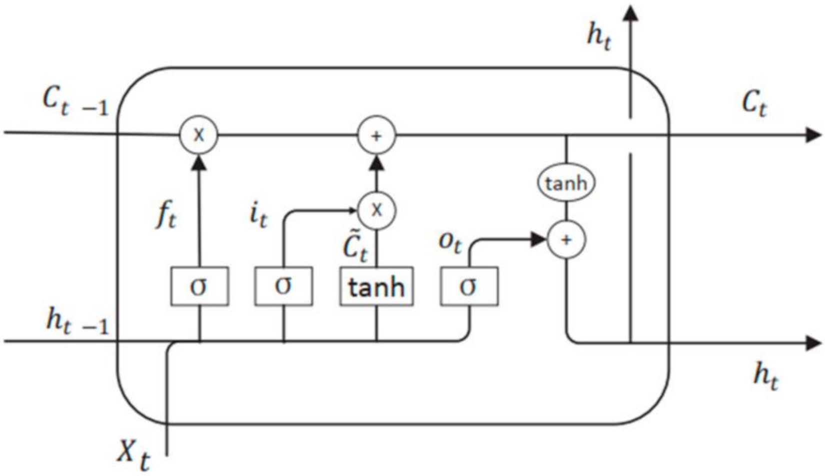

FIGURE 2-2: THE ARCHITECTURE OF A LSTM CELL

The typical architecture of an LSTM cell is made up of gates. In figure 2.2, the three gates are noted as f_t, i_t, and o_t as the output gate. Different functions may activate the gates in the network. The equations for the gates used in ordinary LSTM networks are:

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i)$$
$$f_t = \sigma\left(w_f[h_{t-1}, x_t] + b_f\right)$$
$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o)$$

i_t – input gate

f_t – forget gate

o_t – output gate

σ – the sigmoid function

w_x – the weights for the respective gates' neurons

h_{t-1} – output from the previous LSTM block at timestamp $t-1$

x_t – input at current timestamp

b_x – biases for the respective gate

The input gate interprets the new information stored in the cell state; the forget gate interprets what information is irrelevant and is discarded from the cell state. The output gate is used to provide the activation to the network's final output at time *t*.

In figure 3.2.1, the cell state, candidate cell, and the final output are visualized by the following equations:

$$\tilde{c}_t = \tanh(w_c[h_{t-1}, x_t] + b_c)$$
$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t$$
$$h_t = o_t \odot \tanh\,(c^t)$$

c_t – cell state (memory) at time *t*

$\tilde{c}_t$ – represents candidate for cell state at time *t*

2.5 Dataset

The dataset consists of daily data points on each of the nine variables assumed to impact the spot price of electricity from 01.01.2009 – 31.12.2021. The variables were chosen manually; hence no feature selection was conducted. Furthermore, the selection is characterized by endogenous and exogenous factors that impact the spot price.

Table 1 illustrate our nine variables, a description of units and the data source.

Variable	**Description (units)**	**Data source**
Spot_Price_Average_no	Spot price (NOK/MWh)	Nord Pool
Gas_future_price_EUR	Gas price (EUR/MWh)	EIKON
EUA_future_price_EUR	Price of CO2 emission quotas (EUR/contract)	EIKON
Total_Consumption_Norway_Mwh	Consumption Norway [MWh]	Nord Pool
Net_Exchange_Norway_Mwh	Net power exchange (MWh)	Nord Pool
Water_level_Norway	Reservoir levels (% of capacity)	NVE 1
Downfall_energy_MWh_day_average	Downfall energy (MWh)	NVE 2
Inflow_energy_MWh_day_average	Inflow energy (MWh)	NVE 2
Temperature_oslo	Temperature (Degrees celsius)	Metrologic institute

TABLE 1: ALL INCLUDED VARIABLES IN DATASET

For a descriptive statistics and graphical visualization on each of the variables, read section 7.1 in the appendix. Furthermore, section 3.1 explains how the data was processed.

3. METHODS

3.1 DATA SELECTION

A common problem when working with datasets is missing values. Missing data can lead to lower precision on confidence intervals, weakened statistical power, and biased parameter estimates (Soley-Bori, 2013). There is no perfect way to handle missing values. Ideally, there are no missing values in the dataset at all; however, in the dataset created for this thesis, there is a pattern of missingness on most of the variables. Missingness is mainly because of opening times on marketplaces or because of a lack of recordings. For example, electricity on NordPool is only traded on weekdays, and water levels in Norwegian reservoirs are only recorded weekly.

When working with statistical data material with missing datapoints, the standard practice is to exclude the missing values as default. A limitation of this measure is the large fraction of the original sample that might be excluded. For the dataset used in the analysis, the number of observations would be reduced from over 3400 observations down to 677 if rows with missing data (NaaN) values were to be excluded.

In time-series data like this, a resolving measure is to conducted imputation. Thus, we have used a combination of excluding missing variables and imputation. Dates with missing values on “spot_price_average_no” are excluded. This includes weekends due to closed markets. For the variable “water_level_norway,” imputation is used to handle the limitations of weekly recordings. It is conducted with a Last Observation Carried Forward (LOCF) technique, where missing values are replaced with the value of the last observation. Doing this assumes that the reservoir levels stay constant for a week and change every Monday, even though it fluctuates continuously. A weakness in this technique is that it can introduce a bias to the dataset and make the model perform less efficient when data has a visible trend.

For the variables “downfall_energy” and “inflow_energy,” the weekly data is divided by seven to represent the average downfall and inflow every day for a given week. These are also datapoints that would have fluctuated more have we had real daily recordings.

3.2 LSTM

3.2.1 FRAMEWORK

The LSTM model was constructed and trained with Keras neural network API (Keras.io, 2022). The API is an open-source deep learning library programmed in Python and is backend-supported by Tensorflow. Likewise, Tensorflow is also an open-source machine learning API that supports the framework for the numerical computations (TensorFlow, 2022). When provided with the framework like Keras, the implementation of machine learning in various subject becomes more available.

3.2.2 DATA DIVISION

In the LSTM model, data is divided into three sets. 80 % of the data were used in the training set, and the remaining 20% were split equally into validation and a testing set. The validation set is used to avoid over-fitting to evaluate the ability of the training data. The size of each set was used based on industry standards.

Figure 3-1 visualize the division of the spot-price variable. The color black illustrates the training set, gray the validation data, and turquoise the testing data.

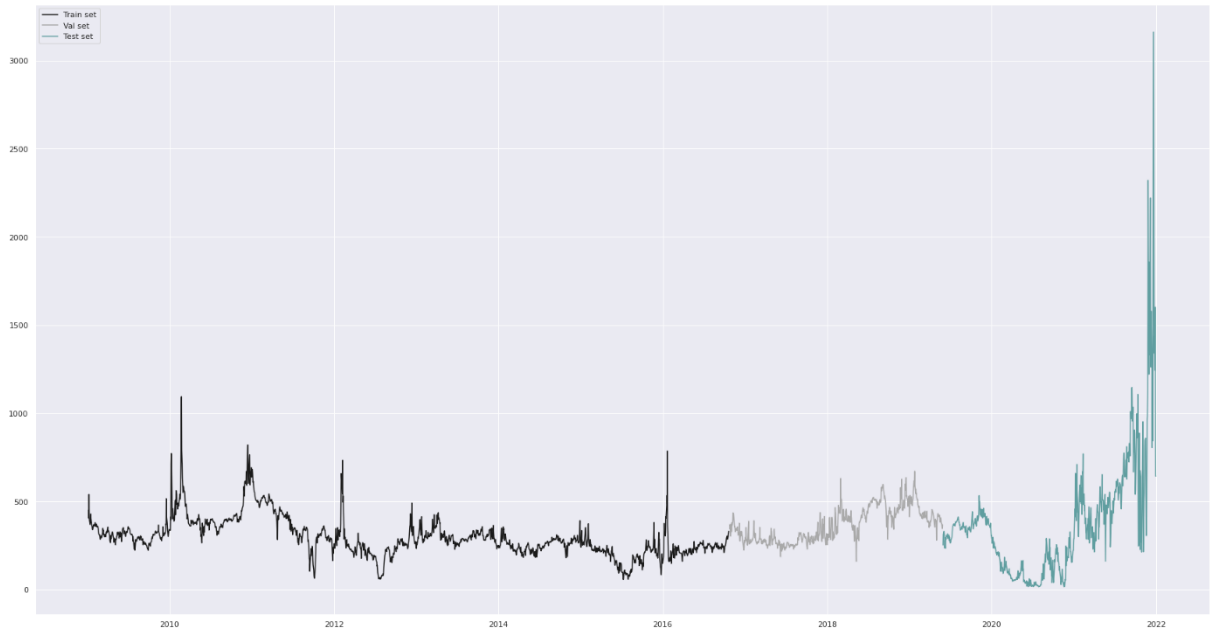

FIGURE 3-1: THE DATA DIVISION ILLUSTRATED IN DIFFERENT COLORS

3.2.3 Data processing

The data was scaled to a value in the interval of [0, 1] in the LSTM model. This is standard practice, as both scaled and normalized data require less GPU to train, thus reducing the time to train. In addition, the prediction seems to be more accurate when using scaled data points. Additionally, it helps when dealing with extreme values. After predicting the model's results, the data was reverted to its original state.

3.2.4 Model architecture

The model performance depends on the hyperparameters and the amount and size of training. The architecture of the network is illustrated in table 2.

Model: "sequential"

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 5, 128)	70656
leaky_re_lu (LeakyReLU)	(None, 5, 128)	0
lstm_1 (LSTM)	(None, 5, 128)	131584
leaky_re_lu_1 (LeakyReLU)	(None, 5, 128)	0
dropout (Dropout)	(None, 5, 128)	0
lstm_2 (LSTM)	(None, 64)	49408
dropout_1 (Dropout)	(None, 64)	0
dense (Dense)	(None, 1)	65

Total params: 251,713

Tranable params: 251,713

Non-trainable params: 0

Table 2: LSTM Network architecture

3.2.5 Input

One day of data points in our dataset is represented by a vector of the size 1x9, where nine is equal to the number of features that concludes our prediction. Therefore, every day’s respective data is placed into a matrix, splitting the days into windows with a fixed length. Next, all vectors are reshaped into a matrix corresponding to a 3D Vector NumPy array to fit the sequential model’s input.

3.2.6 LAYERS

The sequential model groups a linear stack of layers into a Keras model and provides training and inference features on this model (Keras.io, 2022). The model is constructed with stacked LSTM layers and is explained in figure 3-2.

A hidden layer is a general layer between the input- and output layers, where neurons take in sets of weighted inputs. Dense layers contain deeply connected layers; hence, the neurons are connected to every neuron of its preceding layer (Verma, 2021).

Two stacked LSTM-layers are used, and two hidden activation- and dropout layers are used after each hidden layer to prevent over-fitting. The output is formatted to a dense layer with the correct shape, as we only want the predicted spot price as output.

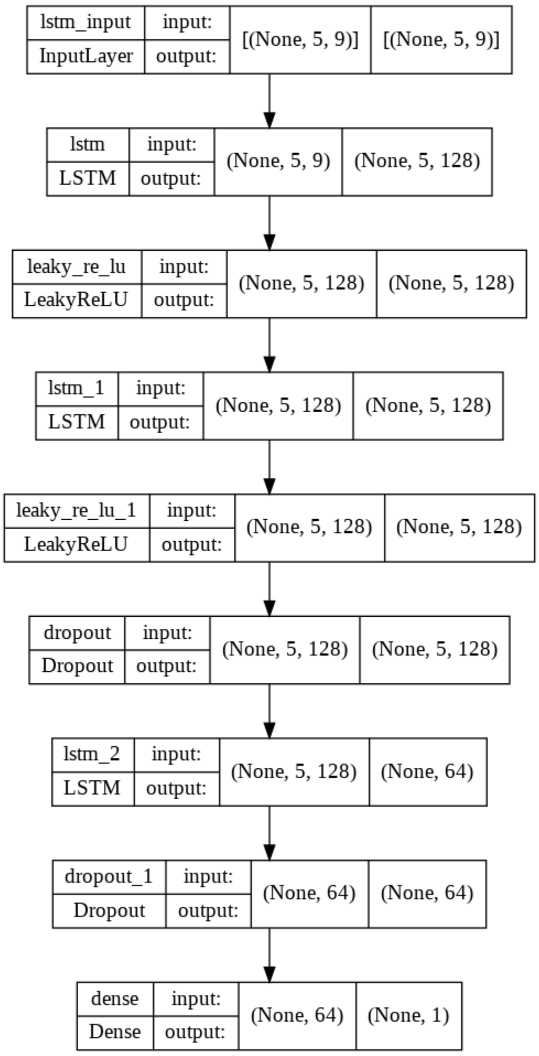

FIGURE 3-2: THE LAYERS THAT MAKE UP THE LSTM MODEL

3.2.7 Early stoppage and epochs

Early stoppage function is also included in the fitting of the model to prevent further over-fitting. The goal is to minimize loss, and with the early stoppage, the metric is monitored. The loss will be checked after every epoch, and if not decreasing, the training stops. After a loss, the function stops further training and does not improve by four iterations. We find the optimal number of epochs to be approximately 40, as shown in figure 3-3. The validation loss flattens out after 40 epochs, and further iteration may cause issues with over-fitting.

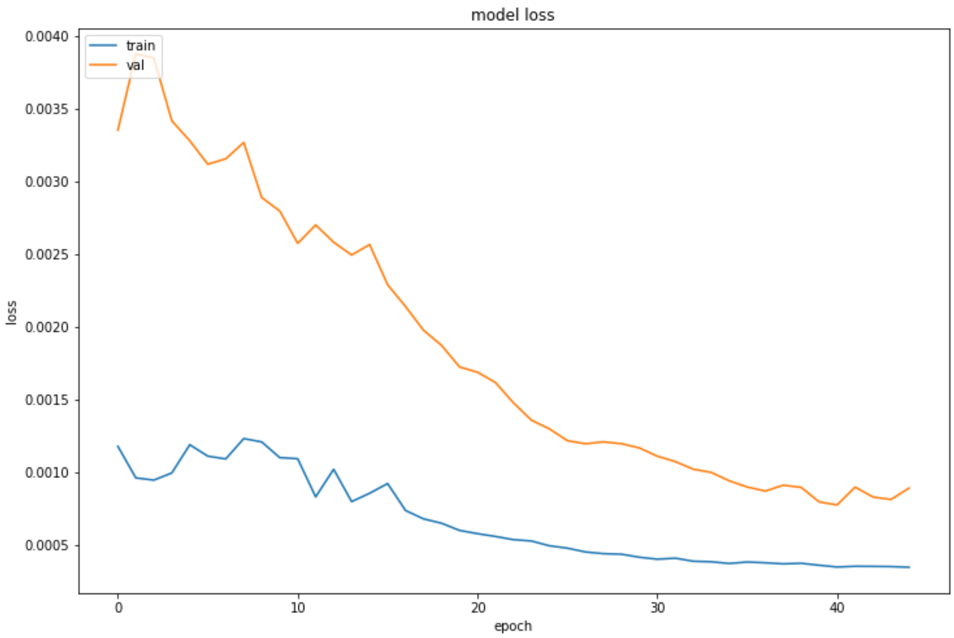

Figure 3-3: Visualization of training loss against validation loss for each given epoch

3.2.8 Loss function and evaluation metrics

During testing, the loss function quantifies the distance between the model's output and desired output to expedite learning. The validation data, whose size is decided manually, is the desired output; we have chosen the validation data to be 20% of the training data. Thus, the training data output is compared to the validation data after each epoch to prevent overfitting by early stoppage during training. Overfitting can occur when training loss is falling whilst validation loss is increasing simultaneously.

Mean squared error was chosen as the loss function. The goal is to minimize the root mean squared error to have our prediction close to the observed values.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

The MAE measures the range of error given the assumption that the errors are normally distributed and not biased.

Mean absolute percentage error is used as the evaluation metric across the two predictive models, and is calculated using the following formula:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i}$$

y_i – actual value

$\hat{y}_i$ – forecasted value

n – total number of observations

3.2.9 OPTIMIZER

Lastly, we used the Adam optimizer when building the model because of its high performance and small memory requirement. It is well suited for large problems in terms of data and parameters (Ba & Kingma, 2014).

3.3 LINEAR REGRESSION MODEL

3.3.1 GOODNESS OF FIT

R^2 tells how much of the total variance in the dependent variable (y) can be explained by the model's independent variables (x) and will range from 0 to 1. The formula for R^2 is:

$$R^2 = \frac{\sum_{i=1}^{N} Ei^2}{\sum_{i=1}^{N} Ti^2}$$

U_i = Residual sum of squares and is the difference between the fitted line and the actual observation.

$$U_i = \sum_{i=1}^{n} (yi - \hat{y}\,i\,)^2$$

E_i = Explained sum of squares and is the difference between the fitted line and the average of the sample.

$$E_i = \sum_{i=1}^{n} (\hat{y}\,i - \bar{y}\,)^2$$

T_i = The total sum of squares is the difference between the observation and the sample average.

$$T_i = \sum_{i=1}^{n} (yi - \bar{y}\,)^2$$

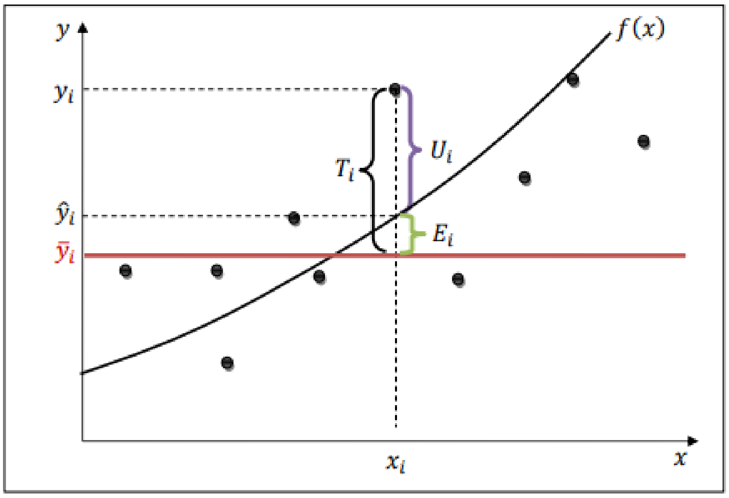

FIGURE 3-4: GRAPHICAL ILLUSTRATION OF THE REGRESSION MODEL AND GOODNESS OF FIT

3.3.2 ASSUMPTIONS AND PROPERTIES TO MULTIPLE LINEAR REGRESSION

In addition to R^2, it is essential to look at the assumptions.

TS.1: Linearity in parameters.

The stochastic process $\{(x_{t1}, x_{t2}, \dots, x_{tk}, y_t): t = 1, 2, \dots, n\}$ follows the linear model:

$$y = \beta_0 + \beta_1 + x_{t1} + \cdots + \beta_k + x_{tk} + u_t$$

, where $\{u_t: t = 1, 2, \dots, n\}$ is the sequence of errors or disturbances. Here, n is the number of observations (time periods) (Wooldridge, 2016).

TS.2: No perfect collinearity.

In the sample (and therefore in the underlying time series process), no independent variable is constant nor a perfect linear combination of the others (Wooldridge, 2016).

TS.3: Zero conditional mean

For each t, the expected value of the error, u_t, given the explanatory variables for all time periods, is zero. Mathematically according to Wooldridge, 2016:

$$E(u_t| X) = 0,$$

TS.4: Homoskedasticity

Conditional on X, the variance of $u_t$is the same for all t:

$$Var(u_t| X) = Var(u_t) = \sigma^2 \text{ (Wooldridge, 2016)}$$

TS.5: No serial correlation

Conditional on X, the errors in two different time periods are uncorrelated:

$$Corr(u_t , u_s| X) = 0$$

Where for all $t \neq s$ (Wooldridge, 2016).

TS.6: Normality

The errors u_t are independent of X and are independently and identically distributed as $N(0, \sigma^2)$ (Wooldridge, 2016).

3.3.3 PREDICTIVE LINEAR REGRESSION MODEL

3.3.3.1 Framework

The predictive linear regression model was constructed and trained with the scikit-learn API library. It features various algorithms to process data. Scikit-learn is straightforward and easy to implement. The model uses the linear regression model and trains the training set before predicting by means of the testing set.

3.3.3.2 Data division

In the linear regression predictor model, data is divided into two sets. 85% of the data were used in training, and the remaining 15% were used for testing.

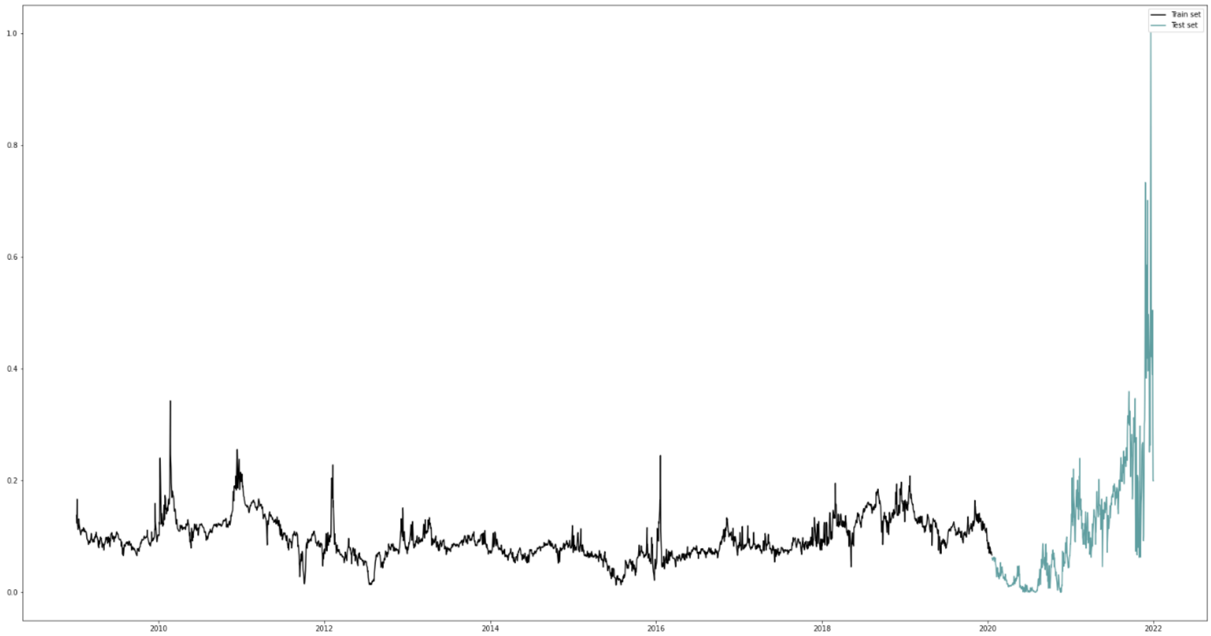

FIGURE 3-5: VISUALIZATION OF TRAINING- AND TESTING SET

Figure 3-5 visualize the division of the spot-price variable. The color black illustrates the training set and turquoise the testing data.

3.3.3.3 Data processing

The data was scaled to a value in the interval of [0, 1] in this model, likewise the LSTM model. Inverse transformation of the scaled data was performed to create a plot, shown in figure 4-1. The reasoning behind this method is found in section 3.2.3.

3.3.3.4 Input

Contrary to the LSTM model, this network does not require the specific vector size as it works like an ordinary multivariate linear regression model. The input is simply the number of independent variables.

3.3.3.5 Evaluation metrics

There is no neural network implemented in the predictive regression model; there is no need for any loss function. The model is evaluated individually by the R^2 score, and the predictions can be evaluated utilizing the root mean squared error, which is explained in section 3.2.8. Also, in section 3.2.8, the standard comparable metric is explained.

3.4 LIMITATIONS

Both models suffer from a lack of both empirical testing and back-testing. There is also a question to be raised concerning the selection of the variables. As LSTM does not have any particular requirements for the input data and suits time-series forecasting. The model does not contain any feature selection and variables could have been weighted in order to compare the linear regression model to machine learning techniques using, i.e., the SHAP package in Python.

Assumptions described in section 3.3.2 may also cause reliability and viability issues. These issues might have been resolved by conducting empirical and specific tests to explore collinearity, homoskedasticity, serial correlation, and normality.

Unwanted bias, namely hindsight bias, can also be introduced and affect the results. Past events seem to be less random and might lead to errors concerning future predictions. Hindsight bias can imply that results are embossed by explanations more than theoretical reasoning and lead to a confusion between noise and signal.

4. RESULTS AND DISCUSSION

4.1 CORRELATIONS AND MULTIPLE LINEAR REGRESSION WITH OLS.

Pearson Correlation matrix

	Spot..	Gas_fut	EUA_fut	Consump	Net_excha	Water_le	Downfall	Inflow_en	Temp
Spot..	1.00								
Gas_fut	0.68	1.00							
EUA_fut	0.50	0.56	1.00						
Consump	0.26	0.13	0.08	1.00					
Net_excha	0.10	-0.09	-0.07	0.22	1.00				
Water_le	-0.16	0.02	0.01	-0.17	-0.45	1.00			
Downfall	-0.10	0.06	-0.02	0.18	0.13	0.19	1.00		
Inflow_en	-0.27	-0.07	-0.01	-0.66	-0.23	0.07	-0.02	1.00	
Temp	-0.24	-0.10	0.03	-0.94	-0.26	0.14	-0.20	0.66	1.00

TABLE 3: CORRELATION MATRIX OF INCLUDED VARIABLES

There is a high correlation between the Spot price of electricity, gas price, and EUA price. Electricity and gas can be seen as substitutes and are therefore naturally correlated. Since EUA works similar to a tax on gas, it affects the demand for gas, shifting the demand over to more climate-friendly alternatives such as electricity. This can be part of the explanation for why these three correlates.

There is also a high correlation between consumption, temperature, and inflow of energy. In the winter, consumption is high, while temperature and inflow are low. This gives some explanation for the seasonality of energy prices. The high correlation between temperature and consumption can cause a multicollinearity problem.

Yearly correlations

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	all
Spot..	1,0	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Gas_fut	0,7	0,209	0,182	0,359	0,462	0,355	0,039	0,545	0,519	0,424	0,789	0,631	0,707	0,659
EUA_fut	-0,2	-0,343	0,692	0,168	0,091	0,229	-0,221	0,229	0,373	0,579	-0,344	0,326	0,686	0,571
Consump	0,6	0,730	0,480	0,781	0,448	0,392	0,800	0,356	0,547	-0,136	0,740	0,529	0,150	0,264
Net_excha	0,2	0,004	0,687	-0,136	0,125	-0,187	-0,235	-0,117	-0,269	-0,415	0,006	-0,010	-0,271	0,096
Water_le	-0,3	-0,090	-0,828	-0,150	-0,371	0,416	-0,192	0,333	0,097	0,312	-0,306	0,183	0,204	-0,159
Downfall	-0,1	-0,105	-0,219	-0,014	-0,339	0,026	0,273	-0,007	0,004	0,017	0,005	0,255	-0,078	-0,104
Inflow_en	-0,4	-0,551	-0,332	-0,600	-0,413	-0,544	-0,767	-0,202	-0,524	-0,269	-0,571	-0,481	-0,204	-0,266
Temp	-0,5	-0,732	-0,437	-0,743	-0,450	-0,417	-0,771	-0,395	-0,564	0,107	-0,673	-0,531	-0,229	-0,242

TABLE 4: YEARLY CORRELATIONS OF INCLUDED VARIABLES

The correlation with European gas prices has gotten more prominent over the last few years. This trend is expected to continue as the Norwegian market gets further interconnected to Europe through the new cables capable of transferring electricity to intercontinental Europe. The new demand from Europe is likely to reduce the correlation between electricity price and consumption in Norway, as seen in 2021.

Multiple linear regression

Number of obs	**3390.00**
F(8, 3381)	188.92
Prob > F	0.00
R-squared	**0.6007**
Root MSE	114.76

spot_price average_no	Coefficient	Robust std. err.	t-value	p-value	std. Beta
constant	359.518	38.2408	44660.00	0,000	
gas_future_price_eur	6.556	0,587	43405.00	0,000	0.476
eua_future_price_eur	4.445	0.2172	20.47	0,000	0.333
total_consumption_norway_mwh	0,000	7.44E-05	-2.63	0.009	-0,080
net_exchange_norway_mwh	0.000242	6.53E-05	25993.00	0,000	0.058
water_level_norway	-81.299	11.7233	-6.93	0,000	-0.086
downfall_energy_mwh_day_average	-0.000089	7.35E-06	-12.06	0,000	-0.086
inflow_energy_mwh_day_average	-0.000087	8.00E-06	-10.84	0,000	-0.144
temperature_oslo	-3.567	0.6933	-5.15	0,000	-0.182

TABLE 5: RESULTS OF THE MULTIPLE LINEAR REGRESSION ANALYSIS

R^2 at 0,6 means that the independent variables explain 60% of the variability in the dependent variable *y* in the model. Dependent variable is “spot_price_average_no”. It is important to note that the assumptions for MLR, listed in 3.3.2, rarely is fulfilled when looking at time-series data as the observations have dependencies among them. Therefore, the following interpretations should mainly be used to find relevant dynamics of the electricity price and map and generalize specific market mechanisms. The MLR was done both in python and Stata and gave the same results.

All variables are significant on a 5% level.

Interpretation:

- *constant.* When all variables are zero, the expected price is 359.5 NOK.
- *"gas_future_price_nok".* If the gas price increases by 1 EUR, the expected electricity price increases by 6,4 NOK.
- "*eua_future_price_nok".* If the price of CO_2 emissions increases by 1 EUR, the expected electricity price increases by 4,4 NOK.
- *"total_consumption_norway*". If the consumption in Norway increases by 1 MWh, the expected electricity price decreases by 0.0002 NOK. The average is 362 892 MWh. Changing measurement levels means a 1 GWh increase in consumption gives an expected decrease in electricity price of 0,2 NOK.
- "*net_export_norway*". If import to Norway increases by 1 MWh, the expected electricity price increases by 0.0002 NOK. The average is -37 177 MWh. Changing measurement levels means a 1 GWh increase in imports gives an expected increase in electricity price of 0,2 NOK.
- "*water_level_norway*". If water levels increase by 1%, the expected electricity price decreases by 0.8 NOK. It is interpreted like this because the variable is in percentage, with values between 0 and 1. If the coefficient is divided by 100, it can be interpreted as a 1% change.
- "*downfall_energy*". If the downfall energy in Norway increases by 1 MWh, the expected electricity price decreases by 0.00008 NOK. The average is 377 355 MWh.
- "*inflow_energy*". If the inflow of energy in Norway increases by 1 MWh, the expected electricity price decreases by 0.00008 NOK, and the average is 380 719 MWh.
- "*temperature_oslo*". If the temperature increases by 1 degree Celsius, the expected electricity price decreases by 3.57 NOK.

Standardized beta is used to compare variables with different levels of measurement and is therefore applicable as we have many variables with different levels of measurement in this model.

Gas- and EUA futures are the variables with the highest std. Beta of 0,476 and 0,333. These are the variables that, according to the MLR, effects the expected electricity price the most.

4.2 PREDICTIVE REGRESSION MODEL

Figure 4-1 visualize the predictive regression models predictive power on the testing data. The orange line illustrates the forecasted electricity price, while the blue is the actual price. Similar to the LSTM model, the model struggle to follow and predict the actual volatile price and does also seem to both under- and over-estimate the fluctuations. The overestimation does not appear to be as heavily as in the LSTM-model and may therefore be more capable to handle extreme values.

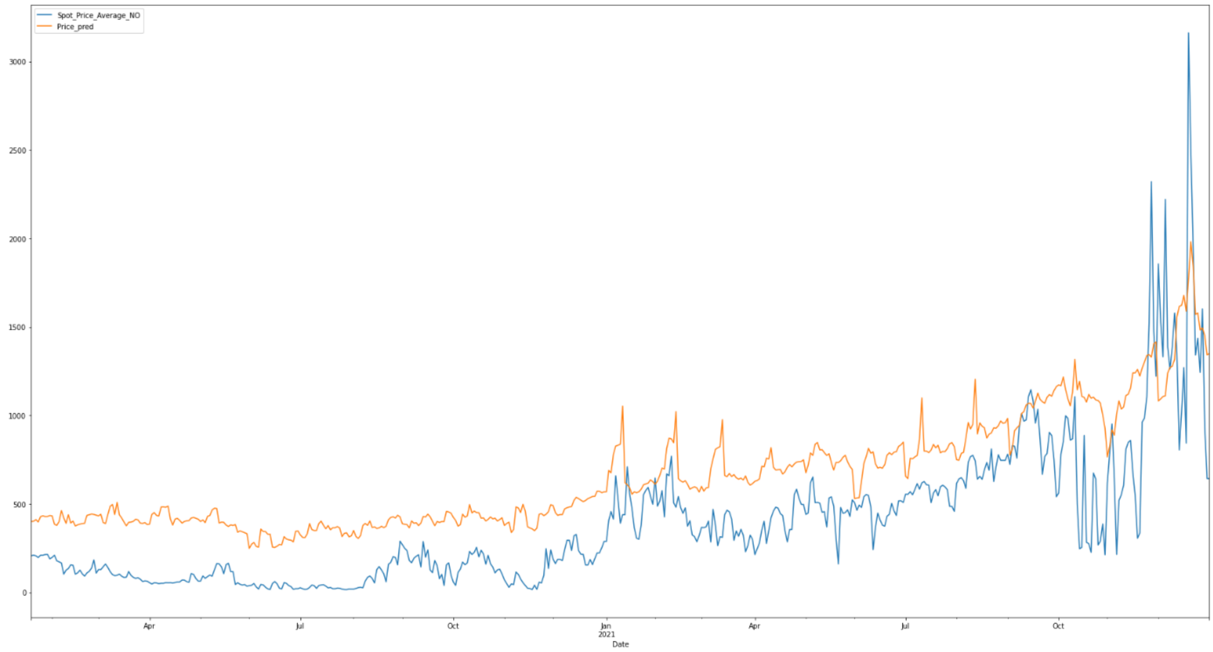

FIGURE 4-1: GRAPHICAL COMPARISON BETWEEN PREDICTED PRICE AND ACTUAL PRICE

The evaluation metric MAPE is 43.95% and interprets that the forecast is off by 43.95% which is slightly worse than the LSTM-forecast. As with the case of the LSTM-model, the regression model is also relatively inaccurate. The reasoning appears to be embossed by the same factors described in section 4.3 but is also most likely affected by some of the assumption mentioned in section 3.3.2.

Figure 4-2 visualize the regression line, which shows the scatter plots' trend line. The target for the model is to minimize the distance of the actual scores from the predicted scores.

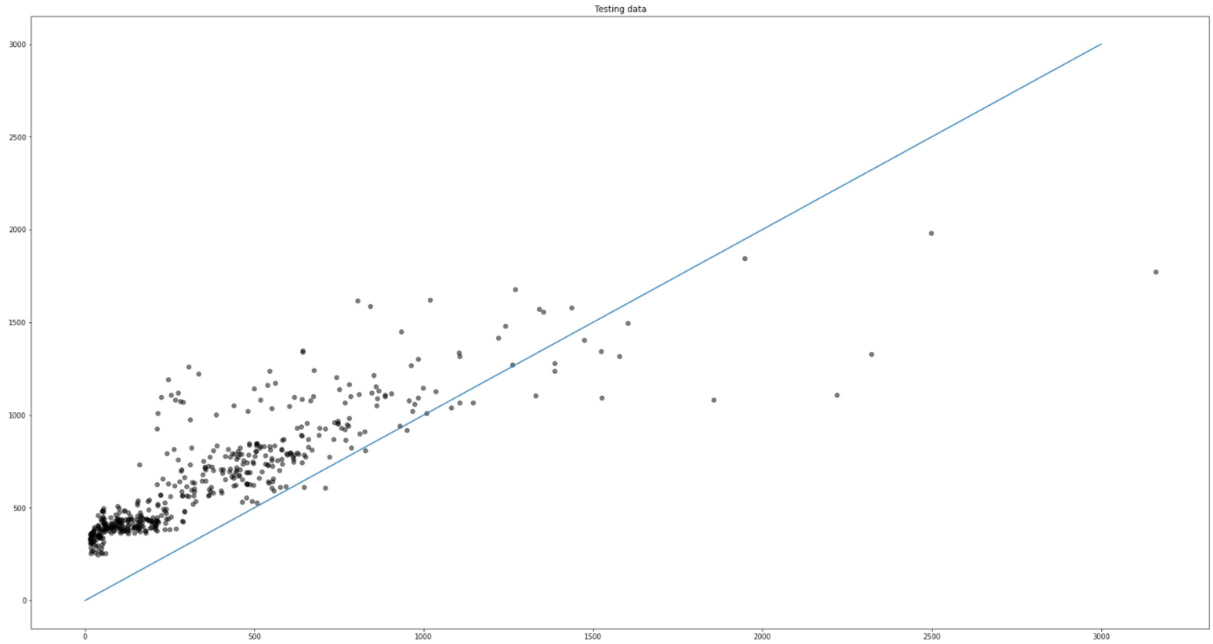

FIGURE 4-2: SCATTERPLOT AND REGRESSION-LINE OF THE TESTING DATA

4.3 LSTM MODEL

Figure 4-2 visualize the LSTM-models predictive power on the testing data. The orange line illustrates the forecasted electricity price, while the blue is the actual price. The models struggle to follow the actual volatile price and seem to both under- and over-estimate extreme fluctuation. Over-estimation is the case for January 2021 and can be explained through the sudden increase in the future contracts of natural gas and CO_2 emission. These two variables had the highest correlation with the electricity price in the dataset.

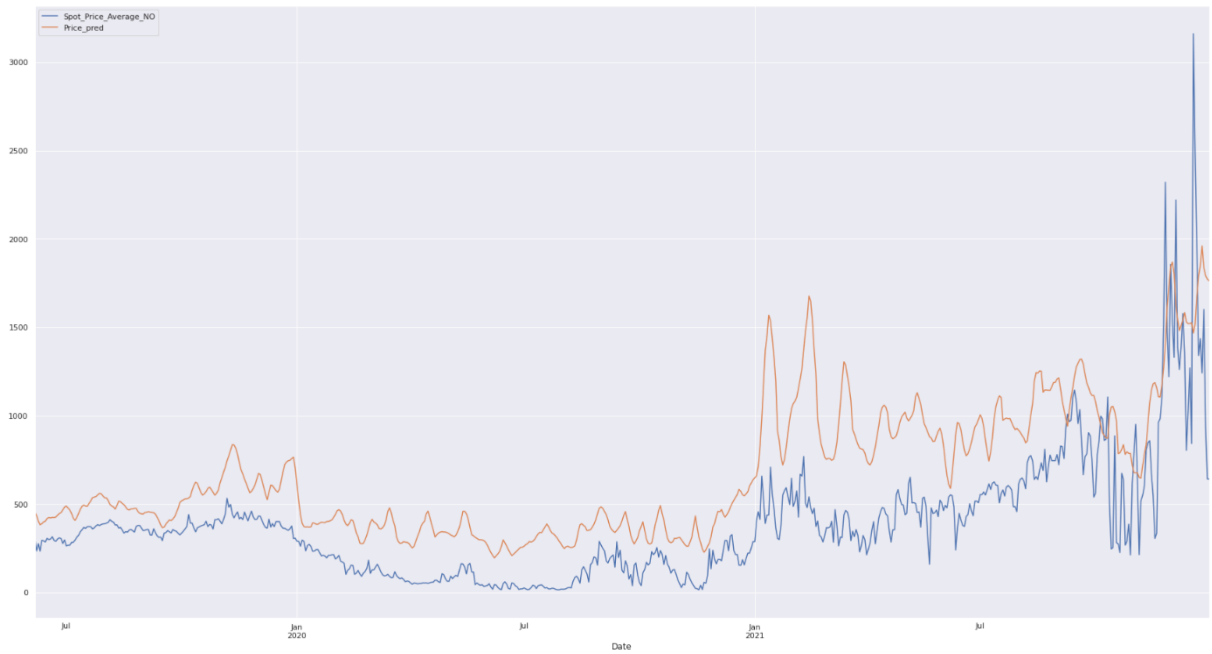

FIGURE 4-3: GRAPHICAL COMPARISON BETWEEN PREDICTED PRICE AND ACTUAL PRICE

The evaluation metric MAPE is estimated at 40.23% and interprets that the forecast is off by 40.23%. This is a relatively inaccurate estimate and might be inflated by specific data points close to zero because MAPE divides the absolute error by the actual data; the values close to zero might significantly inflate the score.

Furthermore, extreme values may also cause high MAPE. The boxplot of figure 4-2 illustrates the values of the electricity price and visualizes how this variable contains extreme values. Extreme values are counteracted by scaling the data, but the MAPE is calculated using rescaled values; this will affect the score contained in this model.

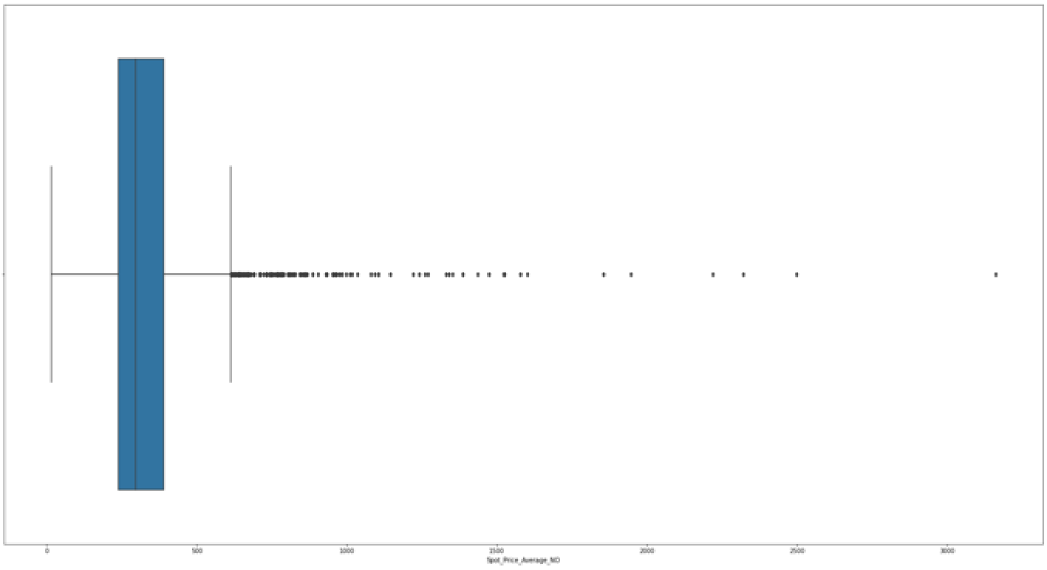

FIGURE 4-4: BOXPLOT ILLUSTRATING EXTREME VALUES FOR PRICE-VARIABLE

4.4 GENERAL DISCUSSION

Electricity in Norway is mainly generated from renewable resources which contributes to the prices' fluctuation. Renewable energy resources are environmentally friendly, and the CO_2 emission from Norwegian electricity production is comparably low with coal-intensive generation (Energifakta Norge, 2021). The volatile pricing reflects the downsides. The generation relies on the weather as water reservoirs need constant water inflow, and low downfall means drained reservoirs that tend to shift the price upwards.

There are notable differences in the electricity prices when comparing seasonal pricing during the summer and winter. Norway relies primarily on electric heating and sees a shift in consumption when outside temperatures drop. The consumption decreases during the summer months, and households use far less energy. Theoretically, we can store surplus energy as downflow- and inflow water fills the reservoirs. Storage is essential as battery storage cannot store this potential surplus energy after generation. Historically, the surplus is used to supply generation to meet increasing demand during the cold winters without seeing a significant increase in the price. In cases where reservoirs receive less participation than expected, the prices are observed to fluctuate and thereby have higher volatility. This seem to be the case as exports have increased due to new cables to the European continental market. Peak demand in the UK can now be facilitated by a surplus in Norway, thus increasing the price inland. Findings support this claim as net exports seem to increase at the same rate that water levels in the reservoirs are decreasing.

Norway is a major supplier to the European continental market (Ministry of Petroleum and Energy and the Norwegian Petroleum Directorate, 2022). As the EU is committed to the Kyoto protocol, which states that the participating countries shall reduce their CO_2 footprint, Norwegian-generated electricity is a sought-after product amongst countries that lack own ability to produce electricity from renewable energy sources. EU Allowances (EUA) permit the emission of one ton of carbon dioxide equivalent (European Energy Exchange, 2020). The EUAs are traded through spot-, future- or option contracts. The contracts are usually a deliverable contract where each Clearing Member with a position open at the cessation of trading for a contract month is obliged to make or take delivery of EUAs to or from a Trading Account within the EUA Delivery Period and under the Rules (ICE Index, 2020). The price of the quotas is directly related to the increase in electricity price.

In addition, Norwegian oil and gas is also a commodity in high demand and a sought-after subject of import in the European continental market. In the global market, Norwegian crude oil and natural gas covered 2020, respectively, 2 and 3 percent of global demand (Ministry of Petroleum and Energy and the Norwegian Petroleum Directorate, 2022). However, Norway supplies 20-25 percent of the EU demand for natural gas, making Norway the third-largest exporter of natural gas. Electricity prices are highly correlated to gas prices, as natural gas might translate into cheaper electricity generation compared with generation from renewable sources.

5. CONCLUSION

Predictive forecasting is essential to maintaining competitiveness in today's electricity market. This thesis has explored the power market through statistical and machine learning approaches to a) analyze and b) forecast the electricity price. Results indicate high correlations between the price of both electricity, gas futures, and EUA futures, and the correlations are seemingly more prominent in recent years. The multiple linear regression model had an R^2 of 60%. Results from the MLR show that the price of Gas- and EUA futures affect the electricity price the most with an std. Beta of 0,476 and 0,333. These results give evidence of increasing harmonization with the Continental European market.

By utilizing the scikit-learn API library in Python, we could predict spot prizes using linear regression. This resulted in a MAPE of 43.95%, meaning that the forecast is 56,05% accurate. Additionally, using the frameworks of Keras and Tensorflow, we were able to construct and train a LSTM model with a MAPE of 40.23%, corresponding to 59,77% accuracy. Performance was similar, but the LSTM yielded slightly better results. Both models struggled to deal with the high volatility and is far from optimal.

Market harmonization may result in less volatility and seasonality over time and increase export, thus, benefiting the Norwegian GDP. On the contrary, there is a concern about whether the increase in electricity price will continue. This can potentially remove the competitive advantage created by Norway's cheap and available energy and incentivize power-intensive industries to either not enter or flee the Norwegian markets. Additionally, it may also permanently change the household economy of Norwegians, as higher electricity bills make for less spending elsewhere.

5.1 FUTURE RESEARCH

Future research conducting predictive price forecasting should establish a robust model that fits the input data. By initially gathering a more comprehensive set of data, the model can benefit from precisely choosing variables through feature selection. This method ensures that the models receive reliable input to improve forecasting results as irrelevant information and data are excluded. For machine learning models, excessive and irrelevant data takes up unnecessary computational power and can wrongly affect data processing in the nodes. Furthermore, as the number of variables increases, the likelihood of randomness disguised as non-randomness affects the results increases. I.e., R^2 increases somewhat for each additionally added explanatory variable.

More extensive back-testing would increase the accuracy of the forecast to minimize MAPE further. Back-testing assesses the viability of the result of a model by discovering how it would perform using broader historical data. Going beyond splitting data into training, testing and validation-set, the model could be tested on a historically broader dataset.

Furthermore, empirical testing should also be performed, as this could further improve the model by choosing optimal hyper-parameters in the model's architecture. I.e., the number of nodes and layers in an artificial neural network model affects the output and could be optimized for increased forecast accuracy. In future research, such methods as k-fold cross-validation or simply testing through trial and error should be considered.

6. BIBLIOGRAPHY

Ba, J., & Kingma, D. (2014). *Adam: A Method for Stochastic Optimization*. Cornell University.

Becker, D. M., & Li, W. (2021, 12 15). Day-ahead electricity price prediction applying hybrid models of LSTM-based deep learning methods and feature selection algorithms under consideration of market coupling. *Energy*(237).

Calin, O. (2020). Deep Learning Architectures: A Mathematical Approach. In O. Calin, *Deep Learning Architectures: A Mathematical Approach* (p. 555). Springer.

CME Group. (2017, 09 25). *CME Group*. Retrieved from Introduction to energy: https://www.cmegroup.com/education/courses/introduction-to-energy/introduction-to-power/understanding-basics-of-the-power-market.html

Energifakta Norge. (2021, February 2). *Ministry of Petroleum and Energy*. Retrieved from Electricity Production: energifaktanorge

European Commission. (2021, 08). *Climate Action*. Retrieved from Emission cap and allowances: https://ec.europa.eu/clima/eu-action/eu-emissions-trading-system-eu-ets/emissions-cap-and-allowances_en

European Energy Exchange. (2020, June 19). *EEX*. Retrieved from EUA Specifications: https://www.eex.com/fileadmin/EEX/Markets/Environmental_markets/Emissions_Spot__Futures___Options/20200619-EUA_specifications_v2.pdf

Fosse, A. L. (2022, 02 08). *Nettavisen*. Retrieved from NVE-sjefen ber nordmenn forberede seg på langvarig strømkrise : https://www.nettavisen.no/okonomi/nve-sjefen-ber-nordmenn-forberede-seg-pa-langvarig-stromkrise/s/12-95-3424242058

Grantham Research Institute on Climate Change and the Enviornment. (2018, July 7). *Climate Change Laws of the World.* Retrieved from The Energy Act, No. 50 of 1990: https://www.climate-laws.org/geographies/norway/laws/the-energy-act-no-50-of-1990

ICE Index. (2020). *ICE Future and Options*. Retrieved from EUA-Futures: https://www.theice.com/products/197/EUA-Futures

Keras.io. (2022). *Deep learning for humasn*. Retrieved from Keras: Simple. Flexible. Powerful.: https://keras.io/

MathWorks. (2022). *MathWorks Deep Learning*. Retrieved from Long Short-Term Memory Networks: https://www.mathworks.com/help/deeplearning/ug/long-short-term-memory-networks.html

Ministry of Petroleum and Energy and the Norwegian Petroleum Directorate. (2021, 08 13). *Norsk Petroleum*. Retrieved from Emission to air: https://www.norskpetroleum.no/en/environment-and-technology/emissions-to-air/

Ministry of Petroleum and Energy and the Norwegian Petroleum Directorate. (2022, March 25). *Norwegian Petroleum*. Retrieved from Exports of oil and gas: https://www.norskpetroleum.no/en/production-and-exports/exports-of-oil-and-gas/

Ministry of Petroleum and Energy. (2019, January 4). *THE LEGAL FRAMEWORK*. Retrieved from Energy Facts Norway: https://energifaktanorge.no/en/regulation-of-the-energy-sector/det-juridiske-rammeverket/#energy-act

Ministry of Petroleum and Energy. (2021, 10 5). *Energifakta Norge*. Retrieved from Norsk Energiforsyning: https://energifaktanorge.no/en/norsk-energiforsyning/kraftmarkedet/#a-market-based-power-system

NordPool Group. (2021, January 1). *NordPool Reports.* Retrieved from 2020 Annual Review: Navigating a Changed World: https://www.nordpoolgroup.com/49eea7/globalassets/download-center/annual-report/annual-review-2020.pdf

NordPool Group. (2022, Jan 01). *NordPool*. Retrieved from Data Portal: https://www.nordpoolgroup.com/en/

NordReg. (2019, 01 10). *Nordic Energy Regulators | NordREG*. Retrieved from An overview of the Nordic Electricity Market: https://www.nordicenergyregulators.org/about-nordreg/an-overview-of-the-nordic-electricity-market/

NVE. (2022). *Kraftsituasjonen. Fjerde kvartal og året 2021.*

Panchotia, R. (2020, August 4). *Medium*. Retrieved from Predivtive Modelling Using Linear Regression: https://medium.com/swlh/predictive-modelling-using-linear-regression-e0e399dc4745

Soley-Bori, M. (2013). *Dealing with missing data: key assumptions and methods for applied analysis.* King's College London.

Tørmoen, A. T. (2022, 02 13). *VG*. Retrieved from Strømprisene: Her er folkets dom: https://www.vg.no/nyheter/innenriks/i/nWlBqB/stroemkrisen-her-er-folkets-dom

TensorFlow. (2022). *TensorFlow*. Retrieved from An end-to-end open source machine learning platform: https://www.tensorflow.org/

U.S. Energy Information Administration. (2021, 12 13). *EIA*. Retrieved from Measruing Electricity - U.S. Energy Information Administration (EIA): https://www.eia.gov/energyexplained/electricity/measuring-electricity.php

Verma, Y. (2021, September 19). *Analytics India Magazine*. Retrieved from A Complete Understanding of Dense Layers in Neural Networks: https://analyticsindiamag.com/a-complete-understanding-of-dense-layers-in-neural-networks/

Wooldridge, J. M. (2016). *Introductory econometrics : a modern approach* (6 ed.). Boston: Cengage Learning, Cop.

7. APPENDIX

7.1 DATASET DESCRIPTION

7.1.1 ELECTRICITY SPOT PRICE

Spot price is the current price in the marketplace at which the electricity can be bought or sold for immediate delivery:

	2009	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**	**all**
count	259,00	261,00	260,00	261,00	261,00	261,00	261,00	261,00	260,00	261,00	261,00	262,00	261,00	**3390,00**
mean	313,60	438,60	374,50	243,50	304,40	254,90	195,10	259,30	281,00	432,30	392,90	120,00	642,50	**327,10**
std	42,00	113,70	114,50	93,80	38,00	34,70	63,70	69,00	36,50	80,60	68,90	80,60	389,00	**181,40**
min	221,60	264,30	63,40	58,70	227,90	179,40	56,60	147,70	186,00	159,00	234,00	15,30	160,00	**15,30**
25 %	289,30	376,20	298,60	200,20	279,90	231,10	151,30	220,80	255,40	384,40	353,60	50,20	425,90	**239,70**
50 %	311,50	397,10	351,00	238,40	298,60	257,80	209,30	241,90	276,90	438,70	385,10	104,00	542,60	**296,40**
75 %	332,10	472,20	469,90	269,60	321,40	278,60	234,00	277,60	296,50	487,50	424,20	181,90	745,70	**389,50**
max	539,10	1093,30	681,10	732,80	435,90	391,00	379,90	784,80	436,20	635,20	670,60	327,10	3161,40	**3161,40**

TABLE 6: DESCRIPTIVE STATISTICS OF SPOT_PRICE_AVERAGE_NO

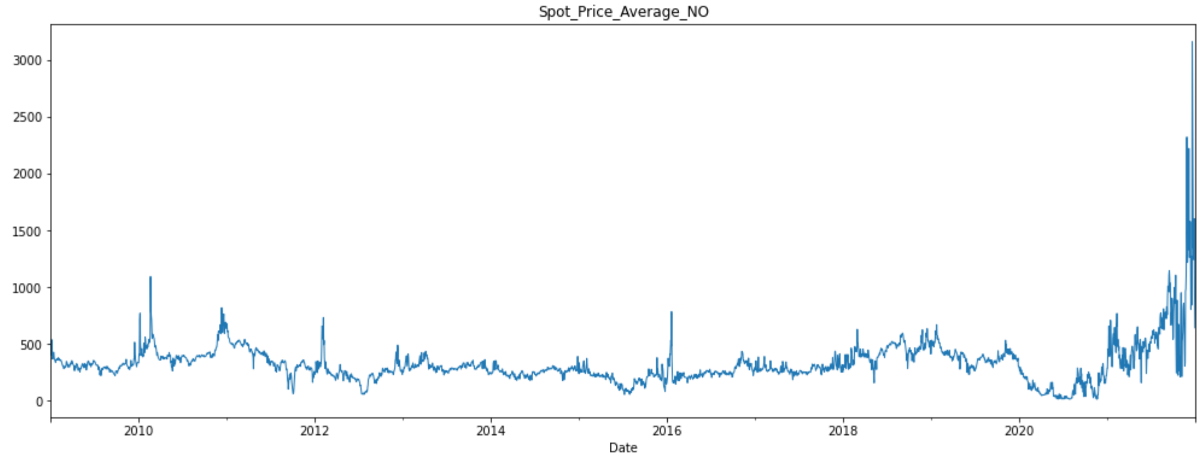

FIGURE 7-1: GRAPHED SPOT_PRICE_AVERAGE_NO

The spot price ranged from NOK 15,3 to NOK 3161,4, with an average of NOK 327. As visualized in the graph (figure 7-1), the highest yearly average occurred in 2021 (average of NOK 642,5) and the lowest in 2020 (NOK 120). Both values are significantly different from the period’s average.

7.1.2 PRICE OF NATURAL GAS FUTURES

This variable contains Prices in EUR of European gas futures. One contract is 10.000 mmBtu which is approximately 2930 MWh.

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	all
count	259,00	261,00	260,00	261,00	261,00	261,00	261,00	261,00	260,00	261,00	261,00	262,00	261,00	**3390,00**
mean	12,00	17,40	22,70	24,90	27,10	20,90	19,80	14,00	17,30	22,90	13,50	9,40	46,90	**20,70**
std	4,30	3,50	1,30	1,90	2,30	2,90	2,10	2,20	2,10	4,60	3,60	3,90	32,70	**13,20**
min	7,10	10,40	16,50	20,00	24,90	14,80	13,90	10,60	14,60	17,40	7,20	3,10	15,80	**3,10**
25 %	9,60	14,20	21,80	23,80	26,00	18,60	18,60	12,30	15,50	20,20	10,40	5,70	20,50	**14,30**
50 %	10,90	18,00	22,70	24,70	26,50	21,10	20,30	13,50	16,60	22,30	13,00	9,30	34,00	**19,50**
75 %	12,00	19,30	23,50	26,20	27,30	23,00	21,10	15,10	19,10	24,60	15,50	12,50	72,40	**23,80**
max	31,80	25,10	25,90	37,80	38,50	26,90	24,50	19,80	23,00	76,00	22,70	19,10	182,30	**182,30**

TABLE 7: DESCRIPTIVE STATISTICS OF GAS_FUTURE_PRICE_EUR

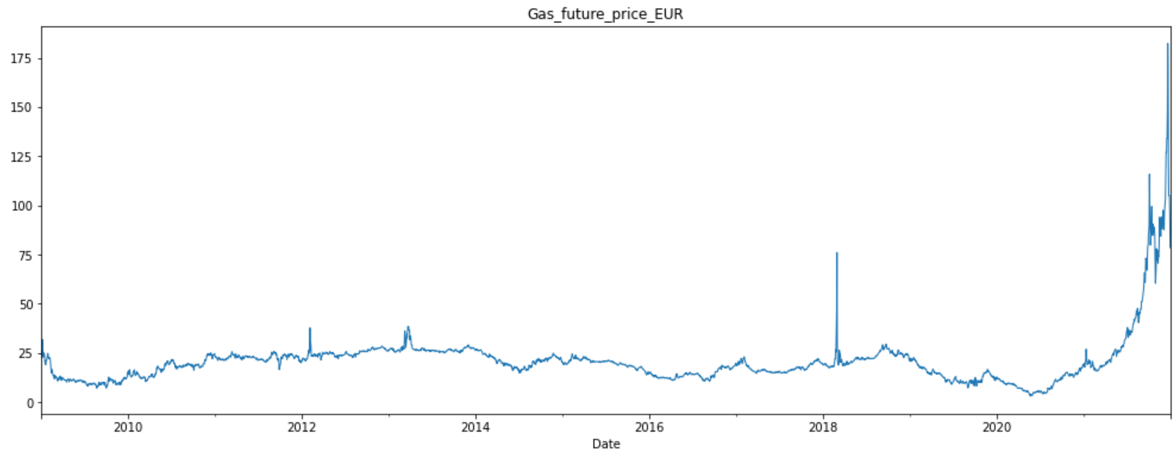

FIGURE 7-2: GRAPHED GAS_ FUTURE_PRICE_EUR

The natural gas price ranged from NOK 3,10 EUR to EUR 183,30. As visualized in the graph and table (figure 7-2, table 7), the highest yearly average occurred in 2021 (average of EUR 46,90) and the lowest in 2020 (EUR 9,40), similar to the spot price of electricity. Both these values also significantly differ from the period's average.

7.1.3 PRICE OF CO2 EMISSION

Prices of EUA futures are measured in EUR. The contract size equals ten metric tons of CO2 emission. Comparably, turbines in petroleum activities in the Norwegian sector emitted 10,28 million tons of CO2 in 2020 (Ministry of Petroleum and Energy and the Norwegian Petroleum Directorate, 2021).

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	all
count	259,00	261,00	260,00	261,00	261,00	261,00	261,00	261,00	260,00	261,00	261,00	262,00	261,00	3390,00
mean	13,80	14,80	13,80	7,90	4,70	6,20	7,80	5,40	5,90	16,20	25,20	25,10	53,90	15,50
std	1,60	0,90	3,20	0,80	0,70	0,70	0,60	0,80	1,10	4,70	2,10	3,70	12,80	13,60
min	8,40	13,00	6,90	6,00	2,90	4,50	6,60	4,00	4,40	7,70	19,20	15,70	31,80	2,90
25 %	13,30	14,00	11,20	7,30	4,40	5,70	7,30	4,80	5,00	13,10	24,00	23,10	43,20	6,50
50 %	14,10	15,00	13,90	7,90	4,70	6,20	7,70	5,30	5,40	16,20	25,40	25,20	53,80	11,10
75 %	14,80	15,60	17,10	8,40	5,10	6,70	8,30	6,00	7,00	20,10	26,60	27,30	60,30	19,30
max	16,60	16,90	18,30	10,30	7,00	7,60	8,80	8,40	8,30	25,80	30,20	33,70	89,40	89,40

TABLE 8: DESCRIPTIVE STATISTICS OF EUA_FUTURE_PRICE_EUR

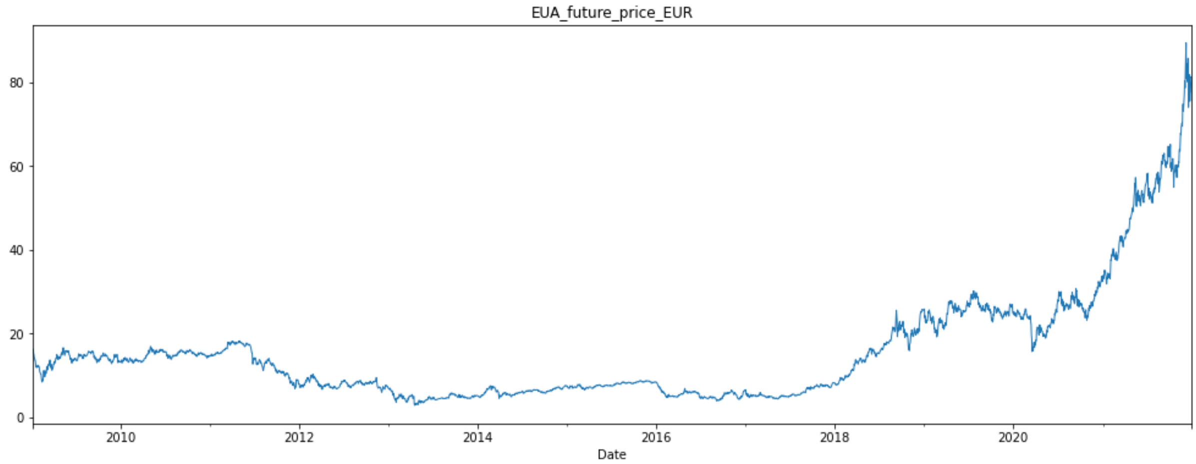

FIGURE 7-3: GRAPHED EUA_FUTURE_PRICE_EUR

The price of CO2 emission seems to have increased steadily in recent years. Regarding the Kyoto protocol, this is affected by the EU regulations but is also expected because the Union-wide cap for stationary installations decreases. Ranged from 2013 to 2020, this is reduced linearly with a factor of 1,74% and has increased to 2,2% as of 2021 (European Commission, 2021).

7.1.4 TOTAL CONSUMPTION

Electricity consumption in Norway in MWh and the table below visualize the Norwegian consumption for the period.

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	all
count	259	261	260	261	261	261	261	261	260	261	261	262	261	3390
mean	336770	366867	343957	356932	357732	351606	358982	368071	371142	377426	372609	368623	386619	362892
std	78378	87218	68948	72483	74202	68596	61496	76900	70585	78731	72564	62129	80678	74667
min	220265	244744	248287	258824	250831	249790	269632	258588	267676	267133	261353	262163	277141	220265
25 %	264735	285883	279329	288301	283275	286162	303819	291458	302386	301318	302228	306867	308567	294512
50 %	328775	348184	336569	352457	350946	346200	352903	368055	370863	361050	368308	369851	376991	357840
75 %	401950	456795	407905	415201	419739	406188	409507	432877	436929	458471	436291	427878	441118	426101
max	484175	535566	482625	518763	526078	502913	493108	542203	514575	530676	518826	481636	559127	559127

TABLE 9: DESCRIPTIVE STATISTICS OF TOTAL_CONSUMPTION_NORWAY_MWH

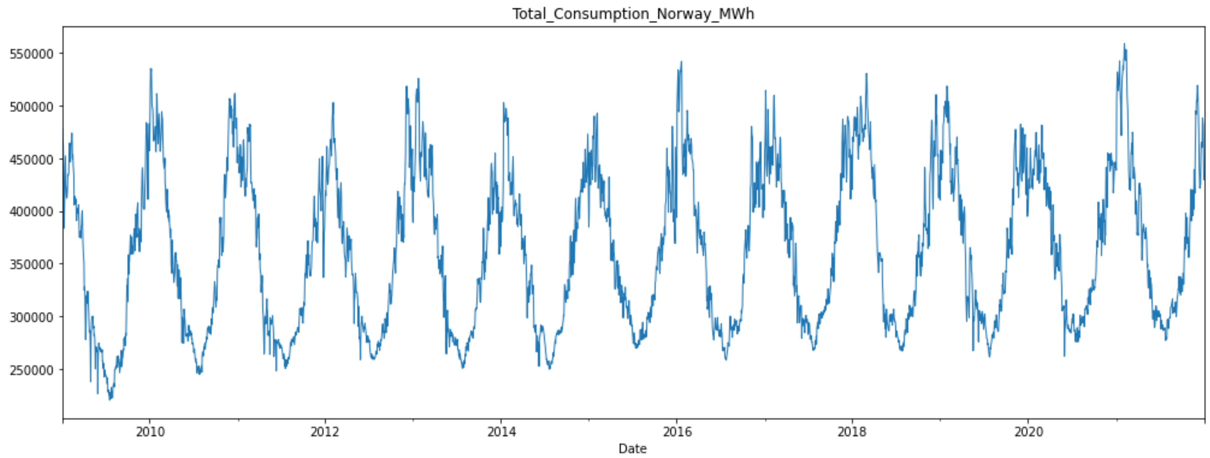

FIGURE 7-4: GRAPHED TOTAL_CONSUMPTION_NORWAY_MWH

Electricity consumption in Norway is stable and is seasonally dependent, thereby stationary, as seen in the graph (figure 7.4). The consumption ranged from 220.265 to 559.127 MWh/day. The highest yearly average occurred in 2021, corresponding to the year with the highest price fluctuations. These values differ somewhat from the period's average of 362.892 MWh/day, but this difference is relatively small compared to other variables.

7.1.5 NET EXCHANGE – IMPORT/EXPORT OF NORWEGIAN ELECTRICITY

Net exchange = (+) import (-) export in MWH. On average, Norway exports more electricity than what is imported. The average net export is 37.177 MWh per day.

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	all
count	259	261	260	261	261	261	261	261	260	261	261	262	261	3390
mean	-32602	11092	-14879	-56040	-23512	-49737	-49765	-53408	-49504	-36738	-11127	-60983	-55929	-37177
std	24520	32042	59399	31812	40322	30196	32413	35744	35541	41613	40884	37818	45190	43703
min	-75508	-70768	-103467	-107034	-90946	-100301	-107377	-111777	-112632	-112771	-105532	-140232	-151626	-151626
25 %	-53229	-12405	-66094	-79074	-53839	-73008	-75304	-78551	-76300	-67484	-41170	-87985	-86635	-70402
50 %	-34777	7647	-33060	-61517	-29104	-56573	-52678	-62354	-60082	-43333	-14944	-67564	-69354	-44106
75 %	-19007	36368	43708	-37731	3722	-28521	-24238	-35665	-30313	-10367	15063	-42351	-29831	-11125
max	67895	86589	101308	94198	85837	38705	75395	84462	64614	118320	100992	72240	120018	120018

TABLE 10: DESCRIPTIVE STATISTICS OF NET_EXCHANGE_NORWAY_MWH

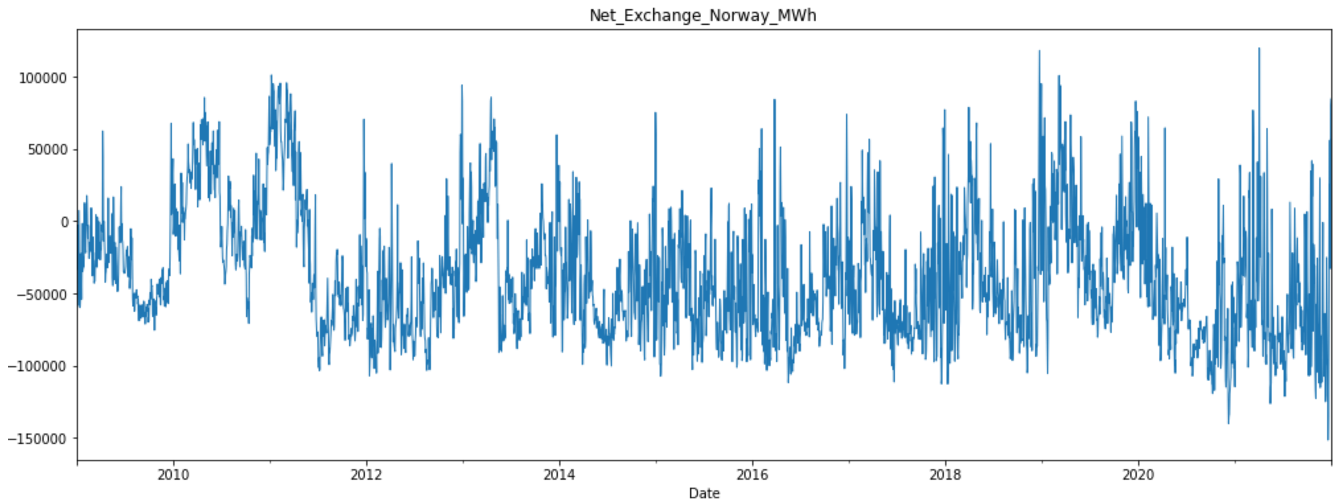

FIGURE 7-5: GRAPHED NET_EXCHANGE_NORWAY_MWH

The import and export of electricity in Norway depend mainly on supply, demand, and transfer capacity. Import typically happens in periods when there is high wind and solar energy production on the continent. Production in Norway can then be reduced, filling up the water reservoirs (NVE, 2022). Suppose the spot price of electricity is higher in other places in Europe, which is typically the case. In that case, it is in the interest of energy producers to export electricity at these higher prices. This will also result in a higher GDP for Norway. Many are concerned that an increasing energy connection to Europe will cause companies to export more to Europe at the expense of Norwegian people and companies. An argument for an increasing connection to the European power grid is that it can reduce volatility in Norwegian spot prices. This is because we can import electricity to keep up with Norwegian demand or regulate water reservoirs. In 2020 and 2021, the export was higher than average and substantially higher than in 2019.

7.1.6 WATER LEVELS IN NORWEGIAN RESERVOIRS

Percentage of water reservoirs filled.

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	all
count	259,00	261,00	260,00	261,00	261,00	261,00	261,00	261,00	260,00	261,00	261,00	262,00	261,00	3390,00
mean	0,620	0,507	0,581	0,701	0,601	0,639	0,645	0,659	0,628	0,583	0,615	0,702	0,602	0,622
std	0,193	0,157	0,260	0,172	0,177	0,172	0,224	0,152	0,196	0,165	0,165	0,226	0,110	0,192
min	0,299	0,226	0,180	0,410	0,249	0,342	0,299	0,361	0,283	0,247	0,313	0,313	0,352	0,180
25 %	0,443	0,374	0,323	0,528	0,459	0,479	0,431	0,562	0,447	0,496	0,456	0,500	0,535	0,458
50 %	0,635	0,541	0,692	0,725	0,686	0,687	0,663	0,692	0,689	0,614	0,650	0,786	0,648	0,655
75 %	0,791	0,654	0,823	0,865	0,753	0,789	0,857	0,790	0,812	0,696	0,745	0,918	0,683	0,785
max	0,892	0,715	0,882	0,922	0,778	0,847	0,928	0,862	0,866	0,813	0,827	0,957	0,822	0,957

TABLE 11: DESCRIPTIVE STATISTICS OF WATER_LEVEL_NORWAY

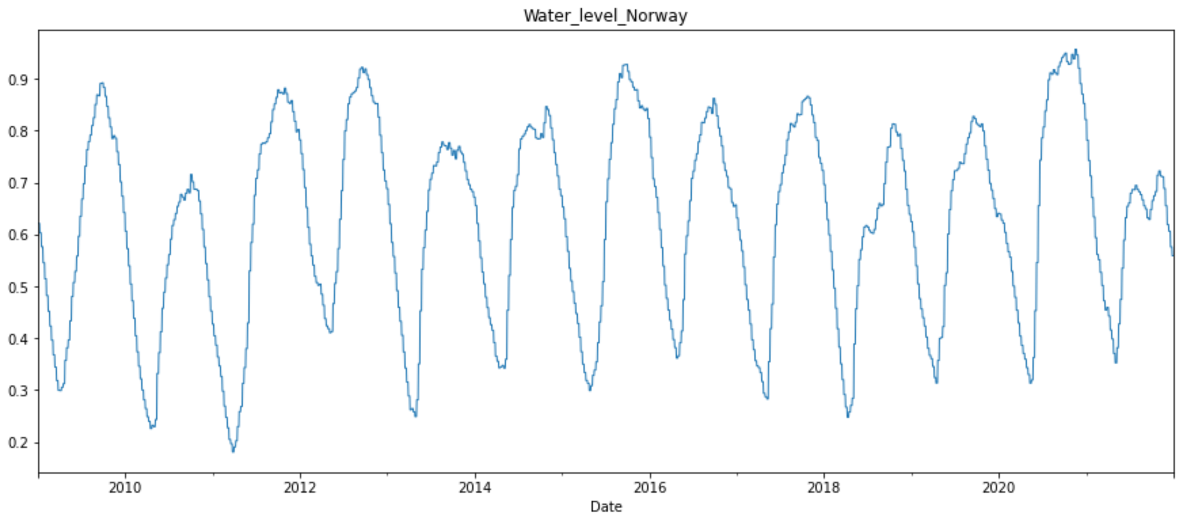

FIGURE 7-6: GRAPHED WATER_LEVEL_NORWAY

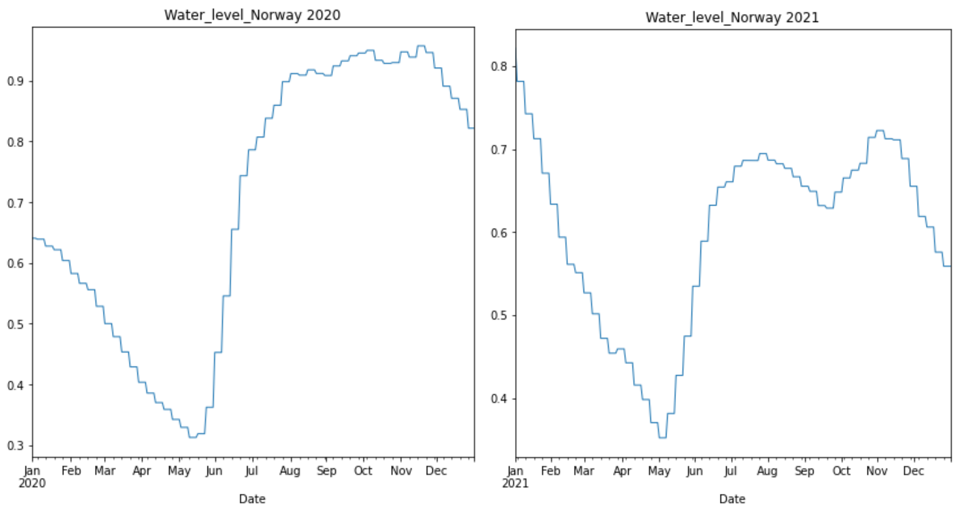

FIGURE 7-7: COMPARISON BETWEEN WATER LEVELS IN 2020 AND 2021

Water levels are also seasonally dependent, generally increasing between May and December and decreasing between December and May. The levels are controlled manually but are affected by the high seasonality of consumption, weather, and other factors. The graph (figure 7-7) shows how water levels were low in the summer of 2021 compared to other years, and one can see a comparison between the water levels in 2020 against 2021.

7.1.7 DOWNFALL ENERGY

Downfall energy in MWh measures the energy stored in rain- and snowfall that either directly or indirectly contributes to water reservoirs via the precipitation field.

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	all
count	259	261	260	261	261	261	261	261	260	261	261	262	261	3390
mean	329756	259082	483143	383477	391807	372464	453541	335725	409773	317757	361559	481566	325724	377355
std	243890	210284	302453	230766	303138	310636	332074	301124	258407	303931	251017	374900	283606	294982
min	-241207	-17496	12559	-101000	-81970	-55077	-140433	-56370	-5519	-241696	-155451	-112749	-245711	-245711
25 %	170446	102299	271332	217271	137556	110254	223534	90199	229588	103944	182686	237691	81150	152114
50 %	296231	188844	442833	404344	333257	283703	402823	269180	405997	323769	366500	452974	291436	340715
75 %	460986	385404	649711	577250	624911	579213	614677	429314	551241	463797	507493	590401	552659	552659
max	827919	1066920	1265469	933021	1160651	1227611	1566049	1191863	967719	1071959	1417963	1573620	876953	1573620

TABLE 12: DESCRIPTIVE STATISTICS OF DOWNFALL_ENERGY_MWH_DAY_AVERAGE

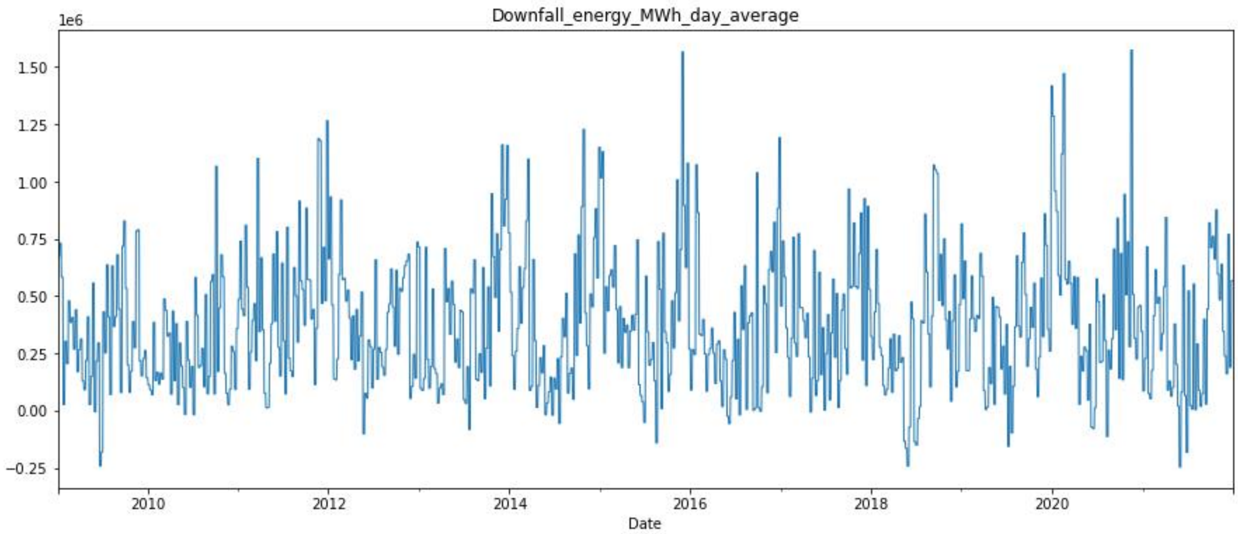

FIGURE 7-8: GRAPHED DOWNFALL_ENERGY_MWH_DAY_AVERAGE

The average downfall energy was 377.355 MWh/day and ranged from -245.711 to 1.573.620 MWh/day in the given period. The highest yearly average occurred in 2020 (average of 481.566 MWh/day) and the lowest in 2021 (-245.711 MWh/day). The graph (figure 7-8) illustrates a cyclic curve with high variations that are not seasonally dependent.

2.1.8 INFLOW ENERGY

Inflow energy in MWh/day measures the inflow of water energy to the reservoirs.

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	all
count	259	261	260	261	261	261	261	261	260	261	261	262	261	3390
mean	358402	279151	441895	395631	361275	404476	434468	349066	390396	370745	337701	489289	336538	380719
std	269838	268571	302897	307474	351926	328156	296179	279073	270991	349295	241943	430933	268519	312785
min	25687	11747	22573	30466	12164	38956	53191	26317	56546	12590	38950	84937	22806	11747
25 %	98210	42691	177468	147760	110831	172214	213397	163271	194342	81533	155241	195516	125355	147204
50 %	381846	228904	423645	326130	277173	311709	346949	262501	322494	277551	284240	358481	266446	317135
75 %	531165	383264	645694	498706	449830	482987	639214	432983	482656	506946	453533	530863	424684	511579
max	1040929	1378081	1466799	1193290	1748607	1481460	1291580	1456769	1135200	1687813	1088686	1985056	1296070	1985056

TABLE 13: DESCRIPTIVE STATISTICS OF INFLOW_ENERGY_MWH_DAY_AVERAGE

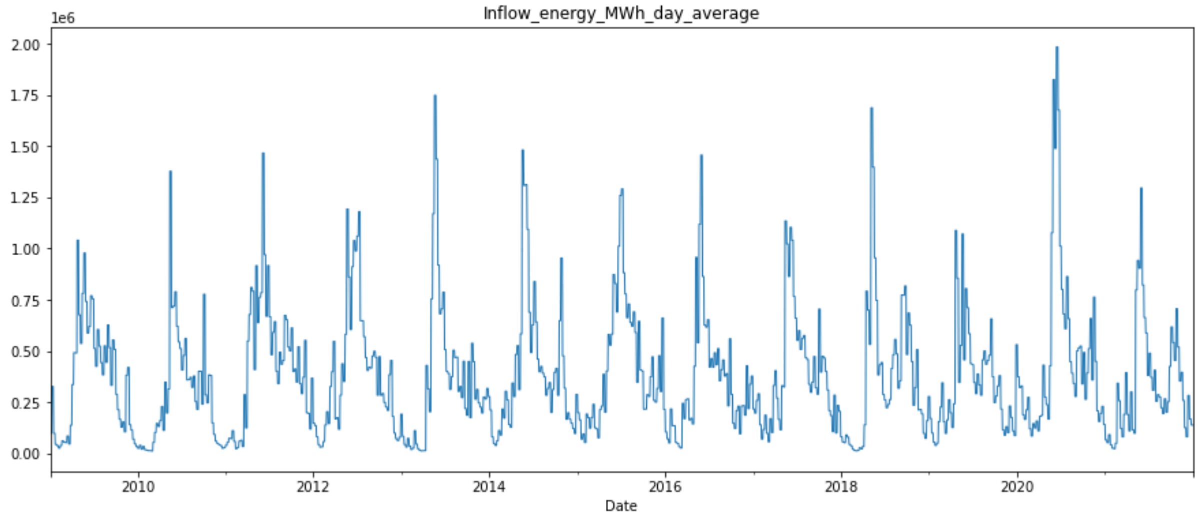

FIGURE 7-9: GRAPHED INFLOW_ENERGY_MWH_DAY_AVERAGE

The average inflow of energy is 380.719 MWh per day, and the graph (figure 7-9) shows the seasonality of the curve. There is usually a high-water inflow after winter (around May) when the snow melts, so water reservoirs start to increase again in May. Also, inflow and downfall energy have very similar yearly averages because they are very close to the same. A lot of the downfall energy will eventually turn into inflow energy.

7.1.9 TEMPERATURE

Maximum temperatures in Oslo on a day-to-day basis, measured in Celsius.

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	all
count	259,0	261,0	260,0	261,0	261,0	261,0	261,0	261,0	260,0	261,0	261,0	262,0	261,0	3390,0
mean	10,9	8,7	11,5	10,4	11,0	12,1	11,8	11,2	10,7	12,1	11,1	12,8	11,3	11,2
std	9,7	10,6	8,9	9,0	9,2	9,6	7,5	9,2	8,1	10,9	8,8	7,7	9,7	9,2
min	-11,8	-13,8	-7,5	-10,6	-9,9	-9,9	-5,1	-11,2	-6,0	-9,1	-7,9	-0,4	-8,5	-13,8
25 %	3,0	-1,0	3,8	3,2	3,7	4,2	5,8	4,2	3,2	1,9	3,4	6,3	4,3	3,9
50 %	11,0	9,7	12,8	11,2	10,2	11,4	11,3	10,1	10,1	12,0	9,9	11,8	11,9	11,2
75 %	18,5	17,9	19,0	18,0	20,2	20,2	18,3	19,8	18,4	21,1	19,1	19,3	20,1	19,2
max	33,0	25,0	28,0	29,8	28,4	33,4	26,5	29,1	25,7	34,6	31,7	29,9	30,2	34,6

TABLE 14: DESCRIPTIVE STATISTICS OF TEMPERATURE_OSLO

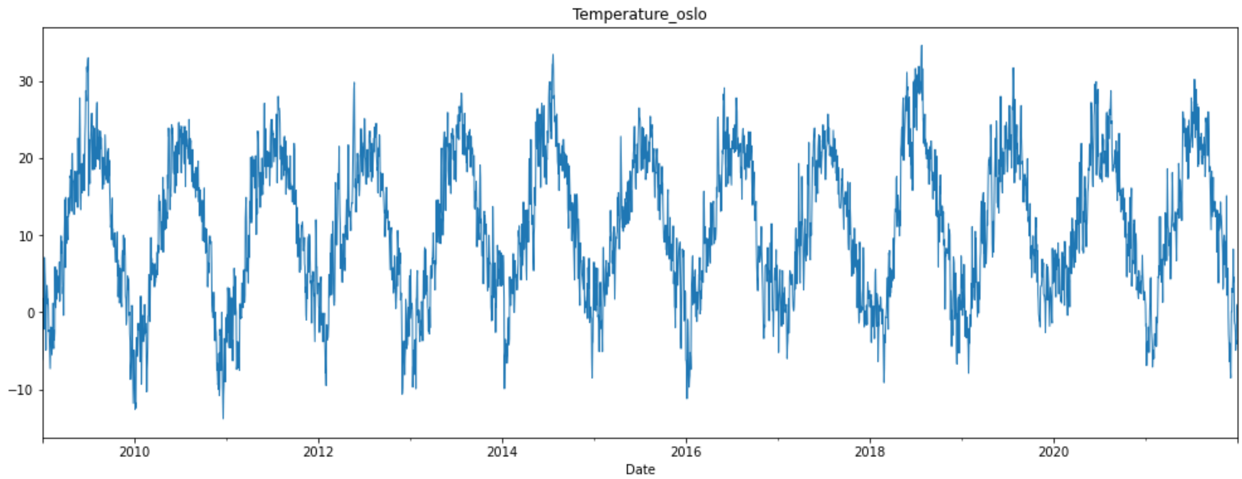

FIGURE 7-10: GRAPHED TEMPERATURE_OSLO

Maximum daily temperatures in Oslo ranges from -13,8 to 34,6 with an average of 11,2 degree Celsius.

7.2 RAW AND PROCESSED DATA

All necessary data is provided.

7.3 PYTHON CODE

Regression model:

https://gist.github.com/thomahje/8dcc58d16dce0b187cafd6d79e02897d

LSTM-model:

https://gist.github.com/thomahje/98e91cc0567c03e3be4fcb2191c95cfa

Further code is provided in zip-file. This attachment is not done public because of privacy concerns on the data provided by NordPool.

zip file is named "Bachelor_files_Johan_Thomas" and is containing:

- "data_variables_csv". This folder contains the CSV files of the variables used to create the dataset. Note that some of the files was cleaned one round before they were saved as CSV files. Complete excel files from NordPool is not provided because of privacy concerns.
- "Datasets_csv". This folder contains CSV files of complete datasets used for linear regression and machine learning. The dataset "all_data_spot_filled" is used for the linear regression using OLS. The dataset "data.csv" is used for predictions.
- "code_files". This folder contains ipynb files of code used for the thesis.
    1. "Bachelor.do" do file from STATA used for MLR using OLS
    2. "Operating_data.ipynb" is the code used to merge data creating the datafiles "consumption_data.csv" and "net_exchange_data.csv"
    3. "DataFrameTotal.ipynb" is the code used to create complete datasets, as well as the code used for linear regression using OLS
    4. "Regression_model.ipynb" is the code use for predicting spot prices using linear prediction.
    5. "LSTM_and_div.ipynb" is the code used for predicting spot prices using LSTM.

The last two files (4 and 5) contain the same content as the github code linked over.

NTNU
Norwegian University of
Science and Technology