LUND UNIVERSITY

SCHOOL OF ECONOMICS AND MANAGEMENT

# Modeling Electricity Prices in the German Energy market - with Applications to Renewables

by

Kaisen Jin, Chibuzo Azuka

May 2022

Master's Programme in Data Analytics and Business Economics

Supervisor: Prof. Joakim Westerlund

# Abstract

Recent focus on the negative effects of climate change has amplified the importance of renewable sources of energy for electricity generation. The contribution of renewables to the energy mix is growing steadily with profound effects on the price of electricity and implications for market participants. In this paper, we employed a similar model to that developed in (Green, 2015) to model the very important shaping vector of the Hourly Price Forward Curve (HPFC) in the German electricity market that is dependent on solar and wind sources of renewable energy. We trained our model using Artificial Neural Networks. However, instead of using price weights as our response variable, we used deviations from the mean to model the shape of the HPFC. We also included calendar information as a variable in the model. We tested the effects of renewables on the shape vector with scenarios of a 15% increase and a 15% decrease in wind and solar generation. Our model indicates that a 15% increase in renewable generation reduces the average price of electricity while a 15% decrease leads to an increase in price. This finding is consistent with the literature and in line with our intuition and proves the existence of a merit order effect. Additionally, we trained a model for short-term price forecasting using a combination of Light Gradient Boosting Machine (LightGBM) and Artificial Neural Networks (ANN) in a two stage forecasting scheme. We used the LightGBM to identify the spike prices and then train both spike prices and normal prices separately using ANN. The predicted spike prices are added to the predicted normal prices. This approach performed better than just training an ANN on the original dataset without separately training the spike prices. We observed that a variable selection using LassoNet did not include both solar and wind generation as important variables for short term normal price prediction, but did include both variables for spike price forecasting.

Keywords: Hourly Price Forward Curves, Power Market, Renewables, Electricity spot prices, Day-ahead market

# Acknowledgements

We wish to express our gratitude to our supervisor Prof Joakim Westerlund from the Department of Economics for his expert guidance and coaching through the writing of this thesis. We also want to thank Rikard Green of Energy Quants Solutions who made this case and provided us with industry insights and data sources. Finally, we acknowledge our family members, friends, classmates, and teachers who, in one way or the other, have made our master's journey exciting.

# Table of Contents

# 1 Introduction

Electricity is a very important commodity. It is present all around us and in everything we do. From education to health to transportation and to agriculture and food production; nothing will ever be the same without electricity. The ever-growing need for more electricity to power different aspects of our life has helped fuel the demand for different types of fossil fuels for electricity production. This has had very devastating effects on our environment. Our climate is changing rapidly with destructive consequences for every part of the world – some more than others. The secretary-general of the UN, Antonio Guterres, in a video message on the launch of the third Intergovernmental Panel on Climate Change (IPCC) report on April 4$^{th}$, 2022 warned that climate change is a global emergency affecting every part of the globe today and reiterated the need for the immediate introduction of renewable sources of energy on a large scale.

This renewed focus on the damaging effects of climate change on our environment and the recent geopolitical tensions in Europe has amplified the focus on the need to increase renewables' contribution to the world's energy mix. In Germany, the percentage of electricity generated through renewable sources has grown steadily in the last 15 to 20 years and currently stands at 47.9%. At the end of the first quarter of 2022, renewable generation was up 20.8% and conventional generation was down 8.1% on 2021 numbers. Solar and Wind are the major sources of renewable energy generation in Germany, jointly contributing 38.9% of Germany's electricity generation in the first quarter of 2022 (SMARD, 2022).

Electricity is a unique commodity in that it is almost impossible to store significant quantities of it. It is also a flow commodity that is produced and delivered over time. Epex Spot (n.d.) presents one of the consequences of this to be that the supply and demand of electricity need to be matched continually to keep the grid stable. Solar and wind generation are dependent on unpredictable weather conditions and are consequently intermittent. This intermittent nature has sometimes caused wide swings in electricity prices depending on whether there are good weather conditions, and therefore high generation, or not. The coupling of most European power markets has made it possible for electricity to flow freely across countries and led to a more efficient matching of suppliers and consumers (SMARD, 2017).

This unique characteristic of electricity and the liberalization of the electricity market in the 1990s has led to an upsurge of interest in research studying factors that affect the price of electricity and how this price responds to changes in these factors. However, many of these studies have used models exploring the impact of all sources of electricity generation including renewables and conventional sources. Only a few of them have explored the effects of different renewable conditions on the price of electricity.

Given that the electricity market is relatively capital intensive and requires upfront inputs that are recouped over time, it is clear that precisely anticipating electricity price trajectory is critical for market players. In this thesis, therefore, our main aim is to put forward models to predict the long-term shape of the HPFC and the short-term (Day-ahead) prices in the German power market. In the short-term prediction, we combine an ANN model with the LightGBM. We employ LightGBM to identify the electricity spike prices and use ANN to train and predict both the spike and normal prices. For the long-term shape prediction, we make use of ANN only because the most recent pricing information is unknown. We want to show the effects of different renewable generation conditions on the accuracy of short- and long-term forecasting and the shape of long-term forecasting. In the end, we hope to provide market participants with a model to predict hourly forward-looking shapes given different renewable conditions.

Our paper is focused on the German market - This market currently serves Germany and Luxembourg. Austria was part of this market until October 2018. We believe that as the largest country and biggest economy in Europe, and also recently a net electricity exporter, Germany is well representative of Europe and the findings of our research can be successfully applied and prove useful in other European markets. We have also limited our research to solar and wind sources of electricity only because these two sources are by far the most significant sources of renewable energy for the German market. As we can see in Figure 1, wind and solar account for the majority of the total renewable generations.

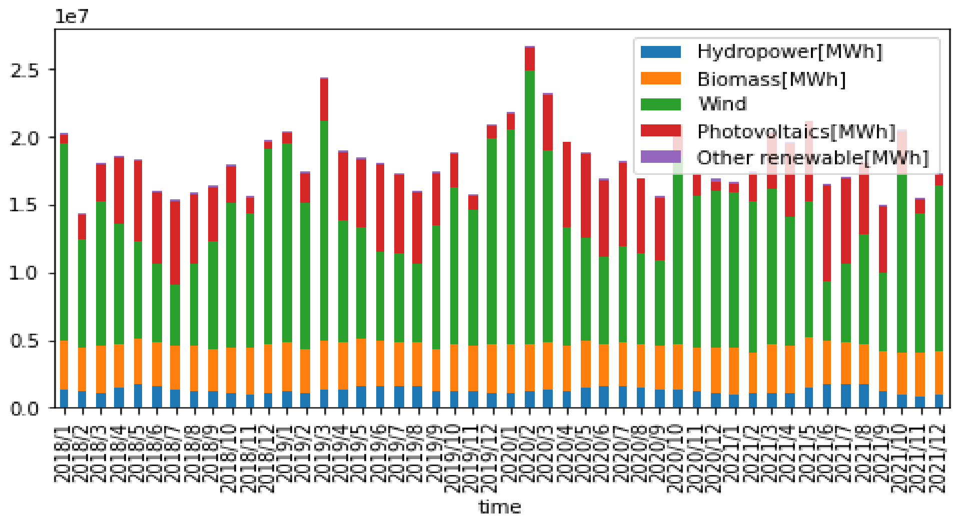


*Figure 1. Gross monthly electricity generation from renewables in Germany from 2018/01 to 2021/12*

Constantly changing relationships between electricity and factors that determine its price means that price and production data of just a few years ago is already unhelpful in predicting future prices. For this reason, we have used only recent data which we consider to still be relevant for our prediction model today. We are using data from 2018 to 2021 in this thesis.

## 1.1 Outline of the Thesis

The thesis is organized as follows. In section 2 we review the literature in our research area including a background of the electricity market in Germany. Section 3 contains key features of our data. In section 4 we present our model and analyze the model results. We present our conclusion in section 5 and suggest areas for further study.

# 2 Literature/Theoretical Review

## 2.1 The Electricity Market in Europe

Before we go into the literature review, we believe an introduction to how the electricity market works in Europe is in order. As we mentioned earlier, a unique characteristic that makes electricity almost impossible to store makes it imperative that whatever quantity of electricity is produced at any one time must be consumed immediately (SMARD, n.d.). Electricity can be bought and sold in either the long-term or short-term market.

In the futures market, futures contracts are used by sellers and purchasers to agree on the quantity and price of electricity to be delivered long in the future (this can be some years) at a price agreed on today. This allows market participants to lock in prices for future delivery of electricity, hedge against future price fluctuations, and helps them with long-term planning and investment decisions. Futures are traded on an exchange like the European Energy Exchange (EEX) which eliminates counterparty risks as the exchange serves as the counterparty to every transaction. Long-term electricity transactions can also happen in over-the-counter (OTC) transactions – these are bilateral contracts consummated outside the exchange. This however exposes the parties to counterparty risks (SMARD, n.d.).

The short-term electricity market is broken into the day-ahead market and the intraday market. Electricity is traded at auctions (held by power exchanges) for next-day deliveries in the day-ahead market. Every day before a set time (noon for most European markets and 11:00 am for Switzerland), traders submit buy and sell bids for every hour of the following day (SMARD, n.d.). The participants indicate the quantity of electricity they are willing to supply or purchase for each price between the minimum and maximum prices of the auction. These bids from different power exchanges are aggregated in a clearinghouse to arrive at the Market-Clearing Price (MCP) for each hour of the following day. The use of a clearinghouse eliminates counterparty risks. The process of arriving at the MCP starts by considering the lowest cost supplier (this is usually the producer or power plant with the lowest marginal cost) and then the next lowest and so on. It also begins by considering the highest bidding consumer and then the next highest and so on. This process continues until all demand is met at a price. This process is known as the merit order because it orders producers from the cheapest to the most expensive and ensures that electricity is produced and supplied by the cheapest producers (SMARD, n.d.). Because renewable energy sources tend to have the lowest marginal cost coupled with their increasing importance, they are said to have merit order effects on prices (SMARD, n.d.).

Every supplier whose offer price is less than or equal to the MCP gets to supply electricity for the hour and receives the MCP while every buyer whose bid price is higher than the MCP receives electricity supply and pays the MCP. This process guarantees that every successful supplier receives an amount that is not less than their offer price while every successful bidder pays an amount that is not higher than their bid price for the hour. Figure 2 shows a graph of the supply and demand for electricity from market participants in the German and Luxembourg power market for the first hour (00 – 01) of May 9, 2022 (for delivery on the 10th of May, 2022). Demand and Supply intersect at the market-clearing price of 216.41 euros per megawatt-hour.

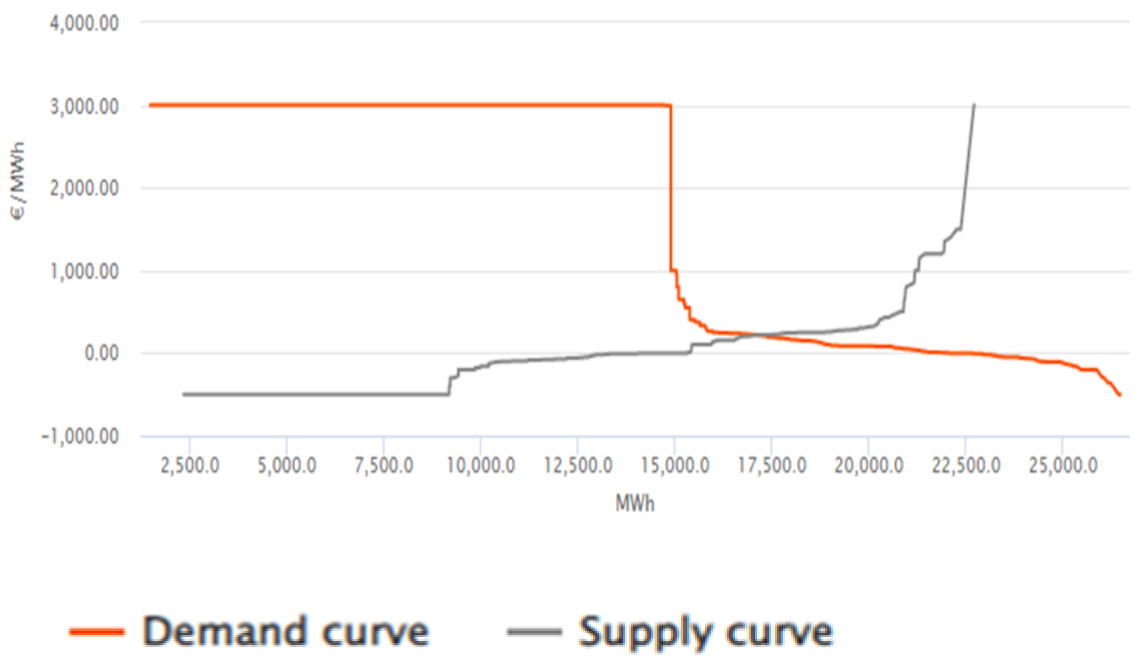


*Fig 2: The determination of the market-clearing price for the first hour (00-01) for 10th May 2022 delivery in the German market. The MCP is determined at the intersection of the aggregate demand and supply orders from market participants on the 9th of May. The MCP is 216 €/MWh. Source: (Epex Spot, n.d.)*

Each day's price profile (shape) would reflect the demand and supply dynamics for every hour of that particular day – for example, prices may fall at night due to low demand and rise during the day as people go about their daily activities and demand increases. Periods of increased supply may also depress prices while the opposite effect may be witnessed during low supply. These price profiles may differ for the seven days in a week, they may also differ depending on the season of the year - that is whether it's winter, autumn, summer, or fall.

The intra-day market is used for transactions of same-day delivery. The European market as a whole is now so integrated and flexible enough that it can accommodate orders for as little as 5, 15, or 30  minutes deliveries depending on the market and trading area involved in the transaction.

Bilateral short-term transactions can also be entered into through OTC transactions with their attendant counterparty risks.

The coupling of the European electricity market has enabled power exchanges to factor in cross-border capacities in the calculation of electricity prices, reducing the price difference between two or more markets and increasing benefits for consumers. (Epex Spot n.d.)

## 2.2 The German Electricity Market

The German electricity market, along with those of other European Union countries, was liberalized in 1998. Before this liberalization, the industry was dominated by 9 regional monopolies that produced and distributed all the electricity within their regions. This monopoly arrangement limited the choices of the consumers within each region to only one provider. It was inefficient and expensive (SMARD, 2017).

Following the liberalization, key aspects of the electricity infrastructure, including generation, operation and distribution infrastructures, were unbundled and opened up to competition and the markets. Today, all aspects of the German electricity market are fully market-driven and the increased competition has led to a more efficient, cheaper and more environmentally friendly electricity industry. This has also led to a dramatic change in the mix of energy sources for the German market with renewable energy sources increasingly contributing a larger and larger percentage (SMARD, 2017).

The passing of the Renewable Energy Sources Act (Erneuerbare Energien Gesetz – EEG) in 2000 paved the way for the increased contribution of renewables to the German energy mix. Germany's renewable energy generation has increased steadily, from about 10% in 2005 to 23% in 2012. It was at 30% in 2016 and as of the first quarter of 2022 renewables contributed 47.9% to the German electricity generation mix, making renewables possibly the most important source of energy in Germany (SMARD, 2017). Figure 3 contains the current breakdown of electricity generation by sources as of the first quarter of 2022.

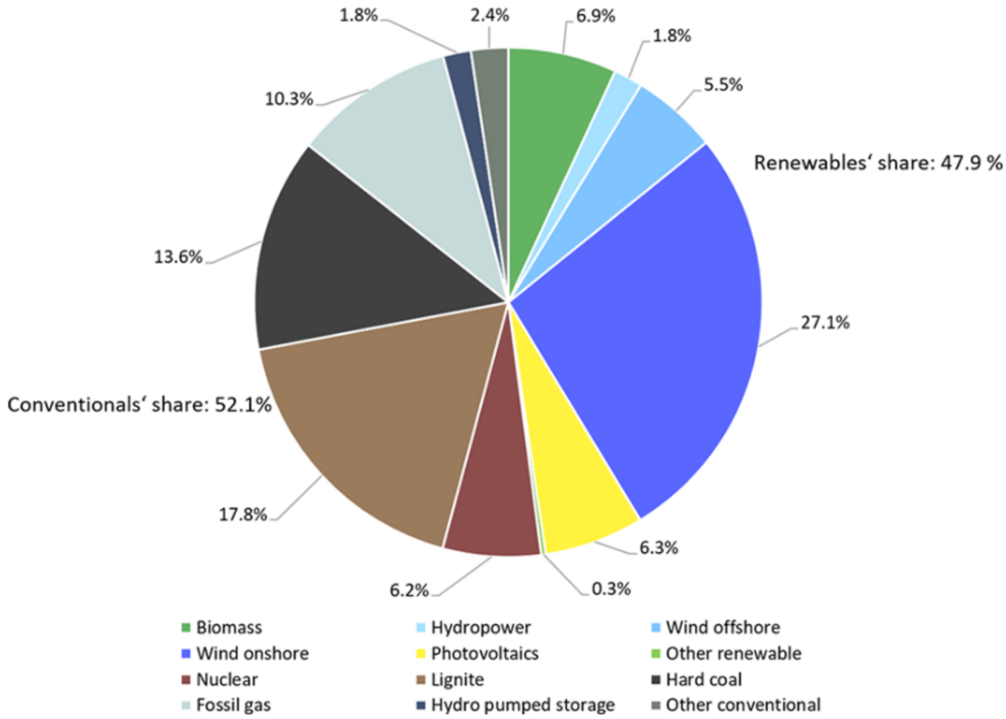


*Figure 3*: *Breakdown of total electricity generation in the first quarter of 2022 by sources. Other renewable include generation from geothermal energy, landfill gas, sewage gas and pit gas. Other conventional include energy from derived gas from coal, mineral oil, waste, oxygen steel furnace gas, blast furnace gas, refinery gas, gas with a high proportion of hydrogen, other by-products of production (for example steel and coke production) and mixtures of more than one fuel type. Source: (SMARD, 2022)*

## 2.3 The HPFC

The HPFC is a very important tool and indicator for energy market participants. In a study by De Jong, Dijken and Enev (2013) for KYOS energy consulting, they noted that the forward prices from the day-ahead market are transformed into a continuous curve of hourly granularity and fed into shaping the so-called HPFC. This is done in two steps; in the first step, an hourly forward-looking vector is estimated and in the second step, this shape vector is calibrated into forward prices in a way that ensures that there is no arbitrage between spot and traded forward prices (Green, 2015). Green (2015) argued in his paper that the HPFC is very valuable for market players because it contains future hourly price patterns and is used by these participants to price contracts more accurately. The shape of the HPFC reflects hourly price differences across the 24 hours in a day revealing daily, weekly, and yearly seasonal patterns. In line with this line of thought, Huisman, Koolen, and Stet (2021) agree that forward power prices contain information about future spot power prices. This is very significant because it means that by having an idea of power prices, traders might be able to predict what future spot power prices will be at a certain time in the future.

## 2.4 Electricity Price Forecasting

Much of the literature on electricity price forecasting is dedicated to the second step in the HPFC determination: the calibration of forward prices themselves with the shaping step not getting enough attention in our opinion. In a key study carried out by Fleten and Lemming (2003), the authors used a quadratic model to construct smooth daily forward curves by combining market prices with forecasts from a bottom-up model. Similarly, Koekebakker and Ollmar (2005) used smoothed data and performed principal component analysis to reveal the structure of the forward price curve, employing a standard lognormal spot price model. While these studies noted that the estimation of seasonal shapes should be done from historical spot prices employing forecasting tools, they were largely still focused on the calibration step and did not provide a strong method for modeling the shape vector.

However, in a remedy to this lack of a strong model for the shaping vector, Green (2015) in his paper developed a shaping model for an hourly forward curve for the Nordics power market, where the profiles depend on the level of the hydrological balance. His model is based on a feed-forward ANN trained on historical hourly electricity spot prices from the Nord pool market. Green (2015) estimated the yearly seasonal effects with historical forward prices from the Nasdaq OMX commodities exchange. He defined the hydrological balance as the amount of available and potential resources used for hydropower production. In order to capture the shape vector, he transformed the spot prices as weights and used these weights as the response variables. His study found that intra-day and intra-week seasonal effects were impacted by changes in the hydrological balance.

In this thesis, we have adopted a similar method to that used by Green. We have modeled a shaping vector for the hourly price forward curve in the German market that depends on the level of renewable energy production in the energy mix. Our model is trained using ANN. We have also developed a short-term spot price forecasting model combining LightGBM and ANN.

Machine Learning is increasingly being employed in the modeling and prediction of electricity prices. However, a range of ML models has been employed by different researchers including long short-term memory, linear regression, convolutional neural networks, Bayesian neural networks, extreme learning machines, and support vector machines. According to Tschora, Pierre, Plantevit, and Robardet (2022) there is currently no consensus reference machine learning model benchmark for electricity price forecasting and model comparison. This is seen as a limitation to good research writing in this field as different research works are not measured by one common standard. Time series based models including seasonal and non-seasonal autoregressive integrated moving average models have been traditionally employed in short-term forecasting but Lago, Marcjasz, Schutter, and

Weron (2021) found that studies comparing machine learning and statistical methods have been contradictory at best.

In a bold attempt to provide a standard reference benchmark for electricity price forecasting, Lago et al. (2021) suggested what they called a state-of-the-art forecasting method and best practice guidelines which includes standards for feature selection and performance measurement. They proposed the adoption of the Lasso Estimated Autoregressive model and the Deep Neural Networks as benchmark models. They trained and tested their models on a set of open-access datasets containing 6 years of price data from 5 different electricity markets around the world. The authors ran their model using the python programming language and documented all their codes and steps in a Github account. They argued that these two models were being suggested because they are easy to reproduce and perform well on price data. The idea is for the industry to adopt their methodology (including the models and their performances on these specific datasets) as the benchmark against which any other model for electricity price forecasting would be measured.

We note that Tschora et al. (2022) adopted the guidelines, put forward in Lago et al. (2021) in their own forecast of electricity prices in three European markets – Germany, France and Belgium using four different Machine Learning models – Convolutional Neural Networks, Deep Neural Networks, Random Forests, and Support Vector Regression over two distinct periods. The authors employed the Diebold & Marino test to compare different models. The Diebold & Marino test is more robust because it compares the loss difference between two model predictions instead of taking an average of the loss over a full dataset. Their results found that different models performed better in different datasets and under different conditions but generally, the Random Forests model performed worse in all situations. They specifically found Random Forests and Convolutional Neural networks to be unsuitable for the price forecasting concept that they studied.

However, given that the study and recommendation by Lago et al. (2021) was only published in July of 2021, there is no evidence in the literature that it has been rigorously examined by others and hence it is yet to be generally accepted and adopted by industry practitioners as a reference benchmark. Because of this, we have not fully adopted these recommendations. Instead, for our long-term shaping model, we have adopted the approach in Green (2015) but modified the response variable as deviations from the average instead of the weights employed by Green (2015). We consider this method easier to understand and communicate. The method is also in line with that presented by Wagner, Ramentol, Schirra and Michaeli (2022). Wagner et al. (2022) adopted the method of spot price transformation in creating a long-term profile for forward curve generation using a Dense Neural Network with an embedding layer to encode calendar information.

Numerous energy and financial companies rely on electricity spot price forecasts, particularly short-term predictions. Price modeling is traditionally made more robust by recognizing and forecasting spike prices independently from the typical time-series spot price prediction. There are also a few publications that focus on using machine learning-algorithms to identify the spike prices with machine learning models such as the ANN used to construct the spike and non-spike prices in order to reduce spike price prediction error and improve the overall performance of short-term spot price forecasts. These procedures are referred to as a two-stage scheme (Shi, Wang, Chen & Ma, 2022).

For the short-term price forecasting, we employed a two stage scheme similar to that presented in Shi et al (2022). The authors employed a Dense Neural Networks framework in the first stage for spike price forecasting and in the second step they employed different variance stabilizing techniques for spike and normal prices to improve electricity price forecasting. In our work, we employ LightGBM which is a novel model developed by Microsoft in 2016. We adopted LightGBM because it employs a leaf-wise split rather than a level-wide split and therefore performs better than most other models in spike price identification. We then train and forecast both the spike prices and the normal prices, separately, using ANN. The output of both is merged together to get our predicted spot prices

According to Janczura, Trück, Weron and Wolff (2013) and Weron (2014), spike prices are broadly classified as, fixed price thresholds, recursive filters, variable price thresholds, fixed price change thresholds, percentage filter, regime-switching classification, and wavelet filtering. Different identifications resulting from different threshold definitions may necessitate the development of different models for capturing price spikes. In this study, we have chosen the percentage filter and have defined spike prices as the top 1% highest prices and the lowest 2.5% smallest prices. These thresholds were chosen subjectively because we consider them to be consistent with the nature of prices in our dataset - there are a lot more negative spike prices than positive ones.

# 3 Data

For this study, we made use of hourly consumption, generation, and day-ahead price data from the German electricity market for 42 months from 2018/01/01 to 2021/12/31. The data is from the German Federal Agency Network which provides various types of electricity related data for Germany and other European countries and is accessed through www.smard.de/en. The dataset contains 30,660 observations and 14 variables. A list of the variables is shown in Table 1

Table 1*:* List of all variables.

| S/n | Variable Name | S/n | Variable Name |
|---|---|---|---|
| 1 | Wind onshore generation | 8 | Nuclear Generation |
| 2 | Wind offshore generation | 9 | Fossil brown coal generation |
| 3 | Photovoltaics (Solar) generation | 10 | Fossil hard coal generation |
| 4 | Biomass generation | 11 | Fosil gas generation |
| 5 | Hydropower generation | 12 | Hydro pumped storage |
| 6 | Other renewables generation | 13 | Other conventional generation |
| 7 | Total grid load | 14 | Residual load |

**Note:** *All the variables are measured in megawatt hours (MWh)*

A review of the price data revealed a very significant increase in the average price of electricity in Germany in the last half of 2021 as shown in Fig 4. To avoid the severe volatility of power spot prices, we adjudged the period after 2021/06/30 to belong to a different data regime and therefore excluded them from our analysis. One reason for this change in regime could be the fact that much of the world economies were beginning to come out from the lockdowns imposed due to the Covid19 pandemic and this led to increased demand for energy (including fossil fuel sources of energy) causing a huge rise in energy prices.

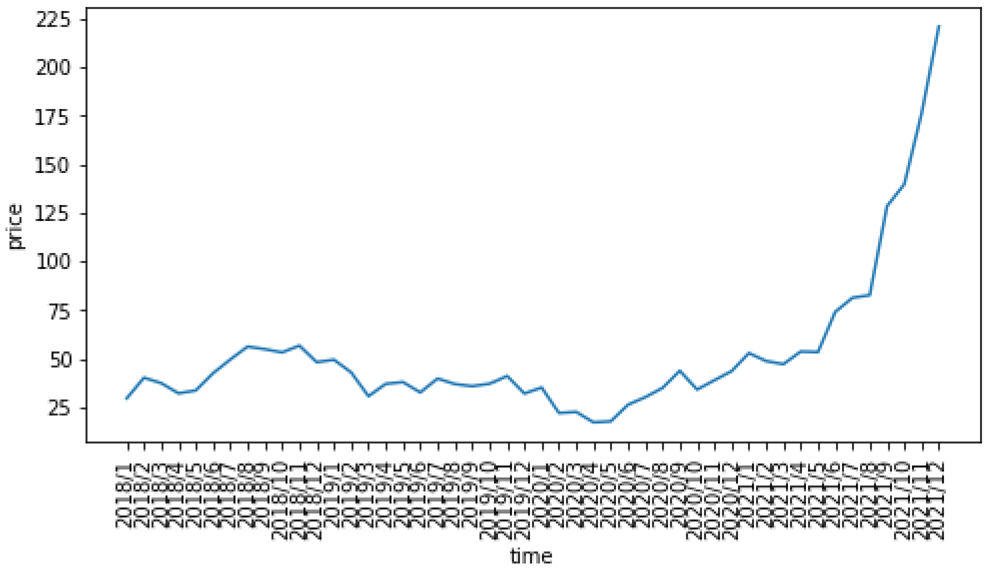


*Figure 4: Average monthly electricity prices in Germany from 2018/01 to 2021/12*

## 3.1 Data Preprocessing

To obtain hourly generation data, all rows in the dataset were grouped and aggregated to an hourly frequency using a time-stamp obtained by concatenating the Date and Time of the day columns. The German prices are calculated by adding the columns Germany/Luxembourg and Germany/Austria/Luxembourg together. The reason is that the single electricity market for Germany/Austria/Luxembourg was divided into two different markets: Germany/Luxembourg and Austria in 2018/10/01 The timestamps of three different tables are then used to merge them together. In subsequent sections, future processes corresponding to each specific model will be described.

# 4 Empirical analysis

As indicated in the Introduction, we will build two applications in this section: short-term and long-term forecasting.

**Short-term forecasting**
In the section on short-term forecasting, the importance is placed on predicting day-ahead hourly prices as accurately as possible. A two-stage model will be used, which is consisting of:

1. A LightGBM classifier will be employed to identify price spikes with deseasonalized prices and supply and demand information. In this stage, an oversampling approach will be used to boost the minority's weight in order to improve classification performance.

2. ANN with a variable selection will be used to forecast the spike price and normal price separately. This procedure is referred to as price calibration.

In the final step, we will compare the results of two-stage models with varying factors to those of non-two-stage models with varying factor combinations.

**Long-term forecasting**
In the section on long-term forecasting, the importance is placed on creating hourly profiles based on historical hourly prices. Instead of the realized price, two new modified time-series will be applied, along with a sinusoidal approach and calendar data to capture seasonality. We will compare the outcomes of the model with renewables to those of models without renewables. Various renewable generation scenarios will be explored as part of the final step.

## 4.1 Short-term forecasting

### 4.1.1 Seasonal decomposition

The spot price for electricity can be viewed as a combination of a stochastic component ($X_t$) and a seasonal component ($S_t$), i.e.

$$p_t = X_t + S_t \tag{1}$$

It has been pointed out by Janczura et al. (2013) that many existing literatures use the stochastic component rather than the real-time spot price data for spike prices identification method. Therefore, We must first discover seasonal patterns $S_t$ to accomplish this.

According to Janczura et al. (2013), the seasonal component $S_t$ is composed of a long-term seasonal component $T_t$ (LTSC) and a short-term seasonal component $s_t$ (STSC), i.e.

$$S_t = T_t + s_t \tag{2}$$

In recent studies, it has been discovered that LTSC removal has a beneficial impact on model performance, despite the fact that LTSC removal is not historically acknowledged as a productive method and was previously believed to be used merely to enhance model complexity (Marcjasz et al. 2017).

As numerous previous studies propose using a sinusoidal function to model periodic patterns (German & Roncoroni, 2006; De Jong, 2005), we could express the long-term component using the sinusoidal Exponentially Weighted Moving Average (sin-EWMA) presented by Janczura et al. (2013).

$$T_t = a_1 sin\left(2\pi(\frac{t}{365} + a_2)\right) + a_3 + a_4 EWMA_t^{0.975} \tag{3}$$

and

$$EWMA_t^a = (1 - a)P_t + aEWMA_{t-1}^a \tag{4}$$

where $a_1, a_2, a_3, a_4$ in equation 3 are derived using a non-linear least square, and $a$ is a decay factor set to 0.975 following the work done by Janczura et al. (2013) and De Jong (2005).

The weekly average deseasonalized price can therefore be used to express the short-term seasonal component,

$$s_t = \frac{1}{7\times24}\sum_{i=1}^{7\times24}\left(S_i - T_i\right) \tag{5}$$

The deseasonalized stochastic component $X_t$ can be stated as the difference between the spot price and the sum of LTSC and STSC.

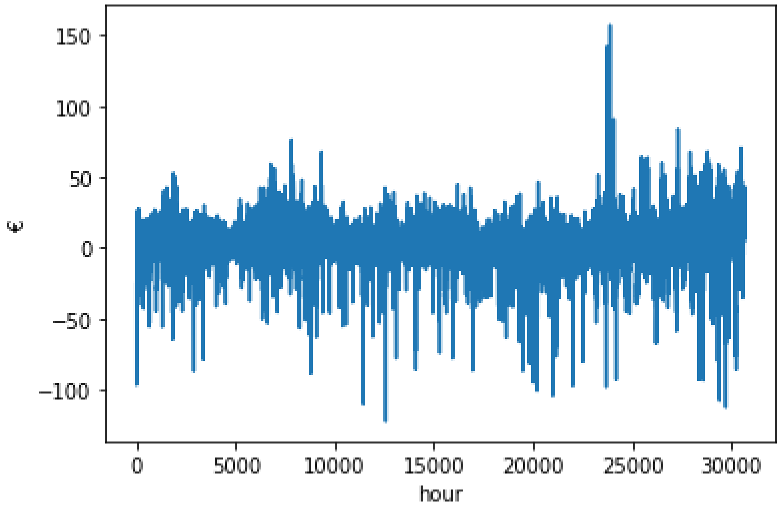


*Fig 5: Deseasonalized electricity spot prices*

## 4.1.2 Spike threshold and further data preprocessing for short-term forecasting

As Janczura et al. (2013) pointed out, most literature chooses the threshold for identifying spikes subjectively. In this particular instance, we would like to define a stringent threshold by choosing a type of variable threshold by combining the standard upper 1 percent prices threshold and Janczura et al. (2013)'s Variable Price Threshold 1 (VPT1) scenario. In the VPT1, 2.5 percent of the highest and lowest prices are considered to be spike prices. Thus, the 1 percent highest and VPT1's 2.5 percent lowest prices are used as the spike price in our scenario. The negative deseasonalized spike prices shown above in figure 5, are more spikey and occur more frequently than the positive spike prices in this particular time period, thus we decided to place more weights on the negative spike prices.

To construct the week and day before lagged variables, the time window of the first week (2018/01/01-2018/01/07) is removed. The variables with a time lag are shown below.

Table 2: Lagged variables

| **Lagged variable** | **Description** |
|---|---|
| $X_{t-24}$ | 24 hour lag variable, deseasonalized price |

| | |
|---|---|
| $P_{t-24}$ | 24 hour lag variable, un-deseasonalized price |
| $P_{t-168}$ | 168 (24×7) hour lag variable, un-deseasonalized price |

The time window of 2018/01/15-2021/01/31 (the first 87.5th of the data) serve as training set, the 2021/02/01-2021/03/03 (the 87.5th-90th of the data) serve as our validation sets, and 2021/03/04-2021/07/02 (the 90th-100th of the data) serve as our test sets, allowing us to limit the number of calculation sources we need to process.

The spike prices are determined using our stated thresholds for the positive and negative spikes on deseasonalized prices. Given the small number of positive and negative spike prices in the sample, in order to improve the classifier's learning ability and increase the sample size, we merged positive and negative spike prices and specified that all prices exceeding the threshold are labeled as category 1 and normal prices as 0. All variables are standardized using z-score standardization.

### 4.1.3 Oversampling

Even when positive and negative spike prices are combined into a single category, they only account for 3.5 percent of the overall data. Since the data has a high degree of imbalance, the classifier will be prevented from gaining sufficient information about a minority category throughout the training process, resulting in the classifier's output being biased towards the majority category. When dealing with imbalanced samples, the model can achieve extraordinarily high accuracy by categorizing all data from minority categories into majority categories; as a result, normal estimates of accuracy cannot be applied to the data in this situation, and we'll go through some more appropriate estimation strategies in the following section. With imbalanced data, industry can increase the minority sample size or lower the majority sample size by oversampling or undersampling, depending on the circumstances. Borderline-Smote 1, a commonly used oversampling technique, which oversample the borderline minority based on its k-nearest neighbors (Han, Wang & Mao, 2005), is used in this case to assist our classifier in capturing the properties of the minority category. To avoid overfitting, we carefully select the two parameters, sampling strategy and m neighbors, depending on the model's performance in the validation set. The model will be presented in section 4.1.4. The number of spikes in the training set was increased from 933 to 1288, and the percentage of oversampled data was increased to 101.32 percent.

### 4.1.4 Identifying the spike price occurrence with LightGBM

Support vector machines (SVM) or Artificial Neural Networks (ANN) are commonly used to predict the occurrence of power spike prices. For the one-hour ahead spike prices forecast, Stathakis, Papadimitriou & Gogas (2021) used an SVM as the target classifier and ANN and XGBoost as the benchmark, while Shi et al. (2021) used ANN as the target model and SVM as the benchmark. Rather than employing the preceding traditional methodologies, we are now introducing LightGBM as our target model. LightGBM has various advantages to many other boosting method like XGboost, including faster training speed, as well as higher accuracy and lower loss by using a leaf-wise (best-first) split approach, as explained by author Erlich_bachman (2017) in the Analytics Vidhya online community.

In our occurrence identification model, the variables we are subjectively considering for classification are deseasonalized lagged prices, $X_{t-24}$, as well as the actual in-feed data Wind offshore, Wind onshore, and Solar, as the same variables were selected by the work done by Stathakis et al. (2021). We also want to improve the performance by sequentially adding other renewable sources from our data, and data from the demand side to see its performance on the validation set with an unconfigured LightGBM. We finally add Hydropower, and Total (grid load), into our model, as they reflect supply and demand data for the energy market. The LightGBM's two key hyperparameters, max depth and num leaves, were set to 5 and 20 respectively by using a 10-fold cross validation grid search. The default value is used for all other parameters.

Commonly, performance is measured with the f-measure for binary unbalanced data that balances both precision and recall rate, rather than accuracy, since a high level of accuracy is achievable with a model that predicts only the majority of categories. The f-measure is defined as:

$$Precision = \frac{TP}{TP+FP} \tag{6}$$

$$Recall = \frac{TP}{TP+FN} \tag{7}$$

$$F - measure = \frac{(1+\beta^2)\times Precision \times Recall}{\beta^2 \times Precision + Recall} \tag{8}$$

Where β is the relative importance of precision vs. recall and often to be set to 1. The f-measure for our validation and test set are 0.61 and 0.56.

### 4.1.5 Replace spike price with normal price

Before estimating the electricity spot price, many literatures would remove the spike price from the spot price and replace it with a less spikey price, i.e. the mean of the normal price. The normal price here is either defined as the threshold value, calculated as the price average before and after the spike, or substituted by the same weekday in a different week in the same month. We used the mean of the average of the prices before and after the identified spike prices, ie, $\frac{1}{2}(P_{t-1} + P_{t+1})$ as a normal price, and if the $P_{t+1}$ are also identified as spike prices, the preceding pricing information, $P_{t-1}$, is supplied at normal prices by following the work done by Weron (2008) and German & Roncoroni (2006). It's normal as our model contains numerous consecutive identified spike prices. We will later refer to these prices as normal prices.

### 4.1.6 Variable selection

The earlier literature focuses on the electricity spot price forecast selecting their features subjectively (Shi et al., 2021) or using a standard technique, such as testing with all conceivable combinations of features with a single-output Gaussian Process Regression (GPR) model (Gabrielli, Wüthrich, Blume, & Sansavini, 2022).

However, we would like to give a more consistent criteria for choosing features when predicting future spot prices, since we are now using ANN to forecast the electricity prices and would like to employ a neural network-based approach with a similar ability to capture non-linearity. Our feature selection model is therefore LassoNet, a recently developed state-of-the-art variable selection method based on the application of Lasso regression to neural networks. Similarly to Lasso regression, LassoNet can arbitrarily establish feature sparsity using a residual feedforward neural network by setting the skip-layer weight for that feature to zero, given a budget. This is not elaborated upon in our work, but it is available for study if interested (Lemhadri, Feng, Abraham & Tibshirani, 2019).

We constructed the LassoNet model separately for the realized price, normal prices, and identified spike prices. To identify a regularization path, the hyperparameters hidden dimension $k$ is manually set to 256 for realized price and normal price from $[2^6, 2^7, ..., 2^{11}]$, and 128 from $[2^3, 2^4, .., 2^9]$ for spike prices, and the dropout layer $d$ is set to 0.2 from $[0,\ 0.1,\ ...,\ 0.9]$ for both realized prices, normal prices and the spike prices, where $k$ and $d$ are chosen based on validation performance, and the default for all other parameters.

The training set for realized prices and normal prices, is the training set that was defined previously, while the training set for spike prices will consist of the identified spike prices on the training set. The ideal combination of $l_1$-Penalty and features are determined by the model's lowest Mean Squared Error (MSE) on the validation set. The model suggests that the combination of fossil brown coal, fossil hard coal, residual load, and undeasonalized price features, $P_{t-24}$, and $P_{t-168}$, yields the lowest MSE for realized prices and normal prices, whereas the combination of biomass, hydropower, wind offshore, wind onshore, photovoltaics, other renewable, nuclear, fossil brown coal, fossil hard coal, fossil gas, hydro pumped storage, other conventional, residual load, and undeasonalized prices, $P_{t-24}$, and $P_{t-168}$, yields the lowest MSE for identified spike prices.
It also proposes a very interesting result, as solar and wind power generation, which were deemed to be two of the most important features for spot prices forecast in the German market due to the renewables' non-negligible share of total electricity generation, were completely

removed from our model for normal price forecasts, and similar results are generated with other different combinations of hidden dimension and dropout layers. This becomes a new question to be studied in the future.

### 4.1.7 Price calibration

This section will discuss the practice of price calibration, which involves forecasting identified price spikes and normal prices separately. Here, only the identified spike price samples from the first stage's output will be entered into the spike price forecasting model. This is the second step of the two-stages model. We will develop three price-calibrated models (two-stage model 1,2,3) with varied input feature combinations, as well as two non-price-calibrated models (original and original+) that forecast the realized spot price only, with varying input features. Models are detailed in Table 3,

Table 3: List of model

| **Name** | **Description** |
|---|---|
| Original | realized spot prices, selected variables |
| Original + | realized spot prices, selected variables + solar & wind |
| Two-stage 1 | spike prices, selected variables - solar & wind<br>normal prices, selected variables |
| Two-stage 2 | spike prices, selected variables<br>normal prices, selected variables + solar & wind |
| Two-stage 3 (Target) | spike prices, selected variables<br>normal prices, selected variables |

A two-layer Artificial Neural Network was utilized to train all models, as it is often used for constructing electricity prices (Lago et al. 2021), with two additional dropout layers to enhance model generalization, and an additional batch normalization layer to enhance our model's performance. Hyper-parameters are selected via Hyperband optimization from the sequence $[2^3, 2^4, ..., 2^8]$ for hidden dimension of spike price, $[2^6, 2^7, ..., 2^{11}]$ for hidden dimension of normal and realized price, and $[0, 0.1, ..., 0.9]$ for dropout layers for all models. These parameters are then configured as follows:

Table 4: Hyper-parameter setting I

| **Model** | **Hidden dimension** | **Dropout** |
|---|---|---|
| spike prices, two-stage 1 | 32/16 | 0.1/0.4 |
| spike prices, two-stage 2, 3 | 32/16 | 0.1/0.2 |
| normal prices, two-stage 2 | 256/256 | 0.2/0.2 |

| | | |
|---|---|---|
| normal prices, two-stage 1, 3 | 2048/128 | 0.3/0.3 |
| realized prices, original | 128/256 | 0.2/0.2 |
| realized prices, original + | 256/256 | 0.2/0.2 |

All models' loss function and activation are ADAM and Rectified Linear Unit (RELU), which are one of the most commonly used by data scientists. All other configurations, like early stopping rules based on the model's performance on the validation set, are the same for all models. We then replace the suspected spike prices in the forecasted normal price with the forecasted value of spike price to derive the forecasted spot prices for two-stage 1, 2, and 3.

### 4.1.8 Result

Using the mean absolute error, we evaluate the performance of our model as follows:

Table 5: Suspected spike price forecasting performance

| **Scenarios** | **Mean Absolute Error** |
|---|---|
| original | 27.50 |
| original + | 26.86 |
| two-stage 1 | 16.50 |
| two-stage 2 and two-stage 3 | 13.87 |

Table 6: Forecasted spot prices performance

| **Scenarios** | **Mean Absolute Error** |
|---|---|
| original | 10.72 |
| original + | 12.85 |
| two-stage 1 | 10.47 |
| two-stage 2 | 11.66 |
| two-stage 3 | 10.40 |

**Note**: *The two-stage models show the results of a normal pricing model in which the normal prices corresponding to the identified spike prices have been replaced with the results of a spike prices model.*

In addition, a Diebold-Mariano test based on the MSE is used to examine the statistical significance of the forecasting accuracy of the benchmark and our model in order to establish the effectiveness of the two-stage treatment, which is shown below,

Table 7: DM-test results

| **Scenarios** | **DM-statistics** | **p-value** |
|---|---|---|

| | | |
|---|---|---|
| original and original + | -14.53 | 2.44e-46 |
| original and two-stage 1 | 5.85 | 5.30e-09 |
| original and two-stage 2 | -0.61 | 0.53 |
| original and two-stage 3 | 5.86 | 5.01e-09 |
| two-stage 1 and two-stage 3 | 3.36 | 7.77e-04 |
| two-stage 2 and two-stage 3 | 19.88 | 7.94e-83 |

**Note:** *In a DM test, a positive value indicates that the model listed first has a higher error than the model listed second while a negative value indicates that the first model has a lower error than the second.*

The slight increase in the prediction results of the two-stage models 1 and 3 relative to the original model is attributable to the slight improvement in the results of the spike price calibration models relative to the original model, as shown in Table 5 and 6. The result of the p-value in the DM-test indicates that the improvement of our two-stage model is statistically significant given a 0.01 cutoff when compared to the original model, as it improves the spike price prediction accuracy. This shows that a two-stage approach would be quite useful when forecasting the short term, particularly the day ahead price forecast.

The MAE results in Table 6 demonstrates that the improvement in prediction outcomes between the original and the original+ is noticeable, as is the improvement between the two-stage 1,3 and the two-stage 2. The statistically significant DM-test statistics and p-value between two-stage 2 and two-stage 3 presented in Table 7 lend credence to the accuracy of the variable selection result. This means that commonly used features such as solar and wind, which are likely to degrade the performance of our model, should not be included in the normal pricing model.

The increase in prediction results of the two-stage model 3 relative to the two-stage model 1 suggests that while the exclusion of solar and wind generation improves the performance of the normal pricing model, excluding them from the spike price value forecasting would decrease the overall performance. Even though the original+ has the worst overall performance, it performs somewhat better in spike price value predictions compared to the original model. This is seen in table 5. The improvement of our two-stage 3 model compared to two-stage 1 is statistically significant as shown in the DM-test, as it increases the accuracy of spike price prediction. This indicates that solar and wind generation play a significant part in the formation of price spikes, which makes sense given that a high volume of solar and wind power is typically associated with extreme low and negative pricing (Smard, 2020), which may also cause negative spikes. Therefore, solar and wind generation should not be excluded from the spike price forecast.

## 4.2 Long-term forecasting

For a very long time, renewables have been regarded as indispensable components for electricity forecasting. In contrast, it was removed from our normal pricing model of short-term prediction in the preceding section. In this section, we are instead constructing a stable long-term electricity forecasting model devoid of two-stage model, as it is hard to forecast spike prices without the most recent pricing information, with the goal of capturing the seasonal patterns of the spot prices and understanding how the supply of renewables, particularly solar and wind power generation, effects predicted hourly profile.

### 4.2.1 Long-term forecasting model

We took a different approach to configuring the seasonality. Since the recent hourly electricity spot prices $P_{t-24}$ and $P_{t-168}$ are not observed, unknown factors would reduce the accuracy of forecasts by influencing the average price level of electricity. In order to stabilize the price level, we no longer use actual spot electricity prices as a measurement for price forecasting, but instead use two indicators called hour to daily deviation (H2D) and hour to month deviation (H2M), which are very similar to the hour-to-month ratio presented by Green (2015), is derived from the deviation of the daily average presented by Wagner et al. (2022). Two indicators are characterized by,

$$H2D_t = p_t - \frac{1}{24}\sum_{i=1}^{24} p_i \tag{9}$$

$$H2M_t = p_t - \frac{1}{|m_t|\times 24}\sum_{i=1}^{|m_t|\times 24} p_i \tag{10}$$

as our response variables, where $|m_t|$ represents the number of days corresponding to that month.

### 4.2.2 Further data preprocessing for long-term forecasting

The time window of 2018/01/01-2019/10/01 (the first 50 percentage of the data) serve as training set, the 2019/10/02-2020/02/06 (the first 50-60 percentage of the data) serve as our validation sets, and 2020/02/07-2021/07/02 (the first 60-100 percentage of the data) serve as our test sets. All variables are standardized using z-score standardization.

The lagging prices $P_{t-24}$, and $P_{t-168}$, are omitted from our sample because the short-term price information is unknown. We also include cosine-sine transformed time variables for the seasonality (Green, 2014; Wagner et al., 2022),

$$m_x(t) = sin(\frac{2\pi t}{12}), m_y(t) = cos(\frac{2\pi t}{12}),$$
$$d_x(t) = sin(\frac{2\pi t}{7}), d_y(t) = cos(\frac{2\pi t}{7}),$$
$$h_x(t) = sin(\frac{2\pi t}{24}), h_y(t) = cos(\frac{2\pi t}{24}), \quad (11)$$

Likewise, we use one-hot to encode the variable 'holiday' to indicate whether it is a German public holiday, where 1 indicates a holiday and 0 indicates a non-holiday.

### 4.2.3 Variable selection

In keeping with the preceding section, a variable selection LassoNet would be employed here as well. The hyperparameters hidden dimension $k$ is manually set to 256 and 512 from $[2^6, 2^7, ..., 2^{11}]$ for H2D and H2M, and the dropout layer $d$ is set to 0.2 from $[0, 0.1, ..., 0.9]$ for both models based on its validation MSE, and the default for the remaining parameters. After adding the cosine-sine transformation and removing the price lagged variables from our model, we observed that all renewables, including our primary interest, solar, offshore wind, and onshore wind, are included. In addition, we are excluding these three variables from our model as a benchmark in order to evaluate the performance of the forecast.

### 4.2.4 Long-term forecasting

Long-term predictions are made using a two-layer ANN with two dropout layers and an additional batch normalization layer. With the Hyperband, two hyper-parameters, the hidden dimensions and the dropout values are selected from $[2^6, 2^7, ..., 2^{11}]$ and $[0, 0.1, ..., 0.9]$,

Table 8: Hyper-parameter setting II

| **Model** | **Hidden dimension** | **Dropout** |
|---|---|---|
| H2D, selected variables | 2048/512 | 0.3/0.3 |
| H2D, selected - solar & wind | 512/1024 | 0.3/0.1 |
| H2M, selected variables | 1024/512 | 0.3/0.4 |
| H2M, selected - solar & wind | 512/2048 | 0.2/0.4 |

The loss function and activation of both models are ADAM and RELU. Other configurations, such as early stopping rules, are identical between the models.

### 4.2.5 Result

The results are measured using MAE and are displayed below,

Table 9: Forecasting performance for selected variables without solar & wind and selected variables using MAE.

| **Scenarios** | **H2D** | **H2M** |
|---|---|---|
| selected variables | 5.32 | 6.66 |
| selected variables - solar & wind | 5.56 | 7.22 |

An additional Diebold-Mariano test based on the MSE is used to examine the statistical significance of the forecasting accuracy of the original scenario and the original scenario excluding the solar and wind components.

Table 10: DM-test for Original without solar & wind and Original

| **Scenarios** | **DM-statistics** | **p-value** |
|---|---|---|
| H2D | 20.72 | 8.03e-94 |
| H2M | 14.91 | 7.17e-50 |

**Note:** *In a DM test, a positive value indicates that the model listed first has a higher error than the model listed second while a negative value indicates that the first model has a lower error than the second.*

The result from Table 9 indicates that the average difference between our model and the actual H2D and H2M deviations is 5.32 and 6.66, respectively, which is a reliable indicator of the actual spot prices over the long term. In addition, unlike the short-term forecast, the inclusion of solar and wind generation sources effectively improves the long-term forecasting performance by capturing the long-term pricing scheme, as demonstrated in the above DM-test from Table 10, which should be emphasized in our study. The positive DM-statistics indicate a better performance by the original model over the original without solar and wind. Given the long-term model, we can assess the effect of renewable energy on the price of electricity.

The following figure 6 and 7 shows the H2D and H2M curve from 2019/10/02/00:00 to 2019/10/08/13:00 in the validation set.

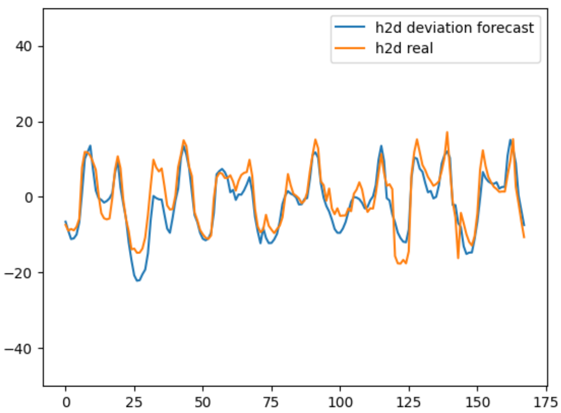


*Fig 6: H2D Deviation*

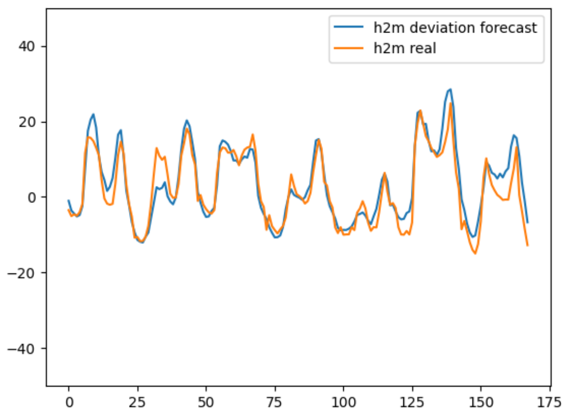


*Fig 7: H2M Deviation*

### 4.2.6 Simple linear model

Prior to advancing to the effect of renewable energy on our long-term forecasting model, we will examine the following simple relation between the change in real spot prices and change in solar and wind power using a simple linear approach that was proposed by De Jong, et al., (2013),

$$\Delta p_h =- \beta_1 \cdot p_{h-24} - \beta_2 \cdot \Delta s_h - \beta_3 \cdot \Delta w_h^{offshore} - \beta_4 \cdot \Delta w_h^{onshore} + \varepsilon_h$$

Where $\Delta p_h$ represents the change in spot prices, $p_{h-24}$ represents 24 hour lagged prices, $\Delta s_h$ represents the change in solar energy generation, and $\Delta w_h$ represents the change in wind power generation. All features except for the response variable, $\Delta p_h$, have been standardized prior to modeling.

Table 11: Linear regression coefficients

| | $p_{h-24}$ | $\Delta s_h$ | $\Delta w_h^{offshore}$ | $\Delta w_h^{onshore}$ |
|---|---|---|---|---|
| β | -4.2851 | -2.2422 | -0.6482 | -8.1483 |
| standard Error | 0.070 | 0.069 | 0.078 | 0.080 |

with a $R^2$ of 0.433.

The output suggests that both change in solar and wind power generation have a significant negative impact on electricity spot prices, as it is in line with our intuition, although renewables would have a non-linear impact on electricity spot prices, and the following step is to construct a model that can capture those non-linear impact. We anticipate, however, that the following section will reach a similar conclusion.

### 4.2.7 Sensitivity analysis

In addition to assessing how the incorporation of renewables affects the forecast performance, as we did in the preceding sections, we were interested in evaluating how solar and wind energy affect the price of electricity with our model.

To be clear, this is a preliminary study that makes no causal inferences; we are only interested in determining whether the change in solar and wind power will affect electricity prices. We refer to this as a sensitivity test.

Consequently, we added two additional scenarios to our long-term forecast: a low wind/solar energy scenario in which solar and wind power generation are each reduced by 15 percent, and a high wind/solar energy scenario in which solar and wind power generation are each increased by 15 percent, in addition to the original scenario.

In the previous section, we used the original model to forecast the h2d and h2m deviations for three different scenarios; the results are shown below.

Table 12: Sensitivity test

| **Model** | **H2D** | **H2M** |
|---|---|---|
| low solar/wind | -0.26 | -1.69 |
| normal solar/wind | -0.31 | -1.79 |
| high solar/wind | -0.35 | -1.89 |

Given our well-trained long-term forecasting model with three distinct scenarios, the result reveals that the model predicts the low renewable scenario would have the highest mean output while the high renewable scenario will have the lowest mean output for both H2D and H2M models, and, the outcomes correspond to our intuition and the merit order effect of electricity prices. This also indicates that our model captures the true patterns of the effect of renewables on energy spot prices to some extent.

# 5 Conclusion

In this thesis, we set out to develop a model for a long-term shaping vector of the hourly price forward curve in the German power market which is dependent on renewable conditions. We limited our considerations to solar and wind sources of renewable generations only. Additionally, we also proposed a model for short-term day-ahead predictions of electricity prices in the German market.

We employed a simple linear regression to estimate the association between changes in renewable generation and changes in electricity prices. Our regression found that solar, onshore wind, and offshore wind generation are all negatively related to electricity prices with onshore wind generation having the biggest coefficient. This is in keeping with general literature and in line with our intuition.

To avoid the impact of unknown factors, we used variations from the daily and monthly means as our response variable. We then went ahead to use this model to predict the shape of the HPFC for a week in October 2019 (October 2 - 8) and found that the model performs quite well especially when using deviations from the Daily mean (H2D) as our response variable. In an attempt to quantify the effects of renewables on electricity prices, we compared the average electricity prices in three scenarios - low renewables, normal, and high renewables. We assumed a situation of a 15% reduction in renewable generation as the low renewable scenario and a 15% increase in renewable generation as the high renewable scenario while the current data is assumed to be a normal scenario. Our model measured the average prices in the three scenarios and, in line with our expectations, we found that a low renewable scenario led to a higher average prices than the normal and high renewable scenarios, with the high renewable scenario having the lowest average prices of the three (results are contained in table 9). We figured that this will enable us to capture the pure effects of renewables. In the model for long-term shape prediction, we excluded the lagged variables because they are unknown for long-term price predictions

In our models, we did not use features that may be currently considered important in electricity price determination like crude oil, coal and natural gas prices for instance. This is in line with the stated objectives of the thesis but also because we foresee a future where renewables are so much more important than they are today. However, given the intermittent nature of renewables, there will always be energy from conventional sources and, for further studies, it might be helpful to incorporate these conventional features into our model to see if our observed relationships and effects change in any way. For the short-term electricity price forecasting, we first employed the LightGBM to identify the spike prices. We then replaced the spike prices with the averages of the two non-spike prices before and after the spike prices to get the normal prices. The normal prices and identified spike prices were trained on an ANN model and used to predict normal prices and spike prices separately. This predicted spike prices is then added to the predicted normal prices to get the predicted spot prices from our model. The result from our model was compared with a result from training the original dataset (without any adjustments) on an ANN and was found to outperform the result from the unaltered dataset. This result is not surprising because we can look at spike prices as deviations which may be more difficult for a single model to capture, hence the two separately helps improve the performance of our model when compared to the model trained on the original data.

A variable selection, conducted using lassonet, for the short-term price prediction produced some interesting results. Both solar and wind generations were found not to be important in predicting short-term normal prices, but were essential in forecasting price spikes. The variable selection for the long-term shape forecasting found all the variables to be important and hence they were all used in our study.

In conclusion, we state that both our long- and short-term models have performed very well in our studied dataset capturing seasonal and non-seasonal patterns in electricity prices in the German market. However, we think that there are areas that should be studied further, for example, the effects of electricity prices and production capacities in neighboring countries on prices in the German market should be investigated. This point is very relevant considering that the electricity market in Europe is currently very integrated. It would also be interesting to study the effects (if any) of electricity prices in Switzerland on the prices in other European countries given that bidding in its day-ahead market closes one hour before those of other countries.

# References

De Jong, C. (2005). The Nature of Power Spikes: A Regime-Switch Approach, *ERIM Report Series Reference*, no. ERS-2005-052-F&A, Available online: ssrn.com/abstract=828173. [Accessed 12 May 2022]

Epex Spot. (n.d.). Basics of the Power Market, Available online: https://www.epexspot.com/en/basicspowermarket [Accessed 12 May 2022]

Erlich_bachman. (2017). Which algorithm takes the crown: LightGBM vs XGBOOST? Available online: www.analyticsvidhya.com/blog/2017/06/which-algorithm-takes-the-crown-light-gbm-vs-xgboost/ [Accessed 12 May 2022]

Fleten, S.E., & Lemming, J. (2003). Constructing Forward Price Curves in Electricity Markets, *Energy Economics*, vol. 25, no. 5, pp.409-424, Available online: https://doi.org/10.1016/S0140-9883(03)00039-2 [Accessed 02 May 2022]

Gabrielli, P., Wüthrich, M., Blume, S. & Sansavini, G. (2022) Data-driven modeling for long-term electricity price forecasting, *Energy*, vol. 244, Part B, pp.123107, Available online: https://doi.org/10.1016/j.energy.2022.123107 [Accessed 12 May 2022]

German, H., & Roncoroni, A. (2006). Understanding the fine structure of electricity prices, *The Journal of Business*, vol. 79, no. 3 , pp. 1225-1261, Available online: https://www.jstor.org/stable/10.1086/500675?seq=1 [Accessed 02 May 2022]

Green, R. (2015). A Power Market Forward Curve with Hydrology Dependence An Approach based on Artificial Neural Networks, The Knut Wicksell Center for Financial Studies, working paper, no. 2015 - 1, Lund University School of Economics and Management

Guterres, A. (2022). Secretary-General's video message on the launch of the third IPCC report, Available online: https://www.un.org/sg/en/content/sg/statement/2022-04-04/secretary-generals-video-message-the-launch-of-the-third-ipcc-report-scroll-down-for-languages [Accessed 16 May 2022]

Han, H., Wang, W. T., & Mao, B. H. (2005). Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning, *Advances in intelligent computing*, part 1, pp. 878-887, Available online: https://link.springer.com/chapter/10.1007/11538059_91 [Accessed 16 May 2022]

Huisman, R., Koolen, D., & Stet, C. (2021). Pricing Forward Contracts in Power Market with Variable Renewable Energy Sources, *Renewable Energy.* Vol. 108, pp.1260-1265, Available online: https://doi.org/10.1016/j.renene.2021.08.086 [Accessed 08 May 2022]

Janczura, J., Trück, S., Weron, R. & Wolff, R.C. (2013). Identifying spikes and seasonal components in electricity spot price data: A guide to robust modeling, *Energy Economics*, vol. 38, pp. 96-110, Available online: https://doi.org/10.1016/j.eneco.2013.03.013 [Accessed 06 May 2022]

De Jong, C., Dicken, H., & Enev, E. (2013). How Renewables Shape the Future in Germany, Available online: https://www.kyos.com/renewables-in-germany/ [Accessed 13 May 2022]

Koekebakker, S., & Ollmar, F. (2005). Forward Curve Dynamics in The Nordic Electricity Market, *Managerial Finance*, vol. 31, no. 6, pp.73-94, Available online: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.194.3209&rep=rep1&type=pdf [Accessed 08 May 2022]

Lago, J., Marcjasz, G., De Schutter, B., & Weron, R. (2021). Forecasting Day-ahead Electricity Prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark, *Applied Energy*, vol. 293, pp.116983, Available online: https://doi.org/10.1016/j.apenergy.2021.116983 [Accessed 10 May 2022]

Lemhadri, I., Feng, R., Abraham, L. & Tibshirani, R. (2019). LassoNet: A Neural Network with Feature Sparsity, preprint, Available online: arxiv.org/abs/1907.12207 [Accessed 13 May 2022]

Marcjasz, G., Uniejewski, B., & Weron, R. (2017). Importance of the long-term seasonal component in day-ahead electricity price forecasting revisited: Neural network models, *HSC Research Reports HSC/17/03, Hugo Steinhaus Center, Wroclaw University of Technology.* Available online: ideas.repec.org/p/wuu/wpaper/hsc1703.html [Accessed 01 May 2022]

Microsoft Corporation. (n.d.) LightGBM: Features, Available online: lightgbm.readthedocs.io/en/latest/Features.html [Accessed 10 May 2022]

Schwartz, E., & Smith, J.E. (2000). Short-term Variations and Long-term Dynamics in Commodity Prices, *Management Science,* vol. 46, no. 7, pp. 893-911, Available online: https://www.jstor.org/stable/2661607?seq=1 [Accessed 01 May 2022]

Shi, W., Wang, Y. F., Chen, Y. Y. & Ma, J. H. (2021). An effective Two-Stage Electricity Price forecasting scheme, *Electric Power Systems Research,* vol. 199, pp. 107416, Available online: https://doi.org/10.1016/j.epsr.2021.107416 [Accessed 30 April 2022]

SMARD. (2017). Electricity market in transition - this is how electricity market 2.0 works, Available online: https://www.smard.de/page/en/topic-article/5892/5972 [Accessed 12 May 2022]

SMARD. (2022). Renewables' share of generation high, Available online: https://www.smard.de/page/en/topic-article/5892/207708 [Accessed 13 May 2022]

SMARD. (2022). Negative electricity prices, Available online: https://www.smard.de/page/en/topic-article/5892/15618 [Accessed 13 May 2022]

SMARD. (n.d.). This is how the electricity market works, Available online: https://www.smard.de/page/en/wiki-article/5884/5840 [Accessed 13 May 2022]

Stathakis, E., Papadimitriou, T., & Gogas, P. (2021). Forecasting Price Spikes in Electricity Markets, *Review of Economic Analysis*, vol.13, no.1, pp. 65-87, Available online: https://doi.org/10.15353/rea.v13i1.1822 [Accessed 05 May 2022]

Tschora, L., Pierre, E., Plantevit, M., & Robardet, C. (2022). Electricity Price Forecasting on The Day-ahead Market Using Machine Learning, *Applied Energy*, vol. 313, p.118752, Available online: https://doi.org/10.1016/j.apenergy.2022.118752 [Accessed 13 May 2022]

Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future, *International Journal of Forecasting*, vol. 30, Issue 4, pp. 1030-108, Available online: https://doi.org/10.1016/j.ijforecast.2014.08.008 [Accessed 03 May 2022]

Wagner, A., Ramentol, E., Schirra, F. & Michaeli, H. (2022) Short- and long-term forecasting of electricity prices using embedding of calendar information in neural networks, *Journal of Commodity Markets*, Available online: https://doi.org/10.1016/j.jcomm.2022.100246 [Accessed 12 May 2022]