Lund University
School of Economics and Management

Master’s Programme in Data Analytics and Business Economics

Modeling German Energy Market Hourly Profiles with a Focus on Variable Renewable Energy

By

Vincent Ball

Thomas Pegoraro

DABN01
24 May 2022
Supervisor: Krzysztof Podgórski

Abstract

This paper investigates the best methods for modeling hourly profiles in the German energy market for the period between 2018 and 2022. Modeling emphasized variable renewable energy (VRE) and included information on the level of energy production, oil price, COVID lockdowns, and historic hourly energy spot prices. Previous research on energy prices has focused on interpretable models; while investigations emphasizing predictive accuracy are sparse and sequestered in industry. This paper is intended to contribute to the understanding of which algorithms and what variables (endogenous and exogenous to the energy market) are best at decreasing the discrepancies between predicted and observed hourly electricity prices.

Four different algorithms were investigated for modeling, linear regression, lasso regression, gradient boosted trees, and a feed forward neural network. Gradient boosted trees accounted for the most variation with an R-squared of 87.7% and promising results on periods of high volatility. Oil price and the share of electricity generated by solar and wind were found to substantially improve predictive accuracy, while COVID lockdowns were less important for prediction. The results from this paper can be used to improve hourly energy price prediction or for comparison by future researchers on different methods.

Acknowledgements:

We would like to express our deepest gratitude to Rikard Green at Energy Quant Solutions Sweden AB for your ideas, industry expertise, and guidance throughout our paper. You have helped us with our research and kept us on the right track. We would also like to thank our supervisor Krzysztof Podgórski, whose valuable advise, and knowledge has been an important factor to this paper. Additionally, we would like to thank Joakim Westerlund for his help and assuring us we are on the correct path. Finally, we thank our friends and family for listening to us, and supporting us so we could stay focused along the way. Thank you all so much!

Section 1: Introduction

Electricity is a unique commodity that experiences volatility in supply and demand patterns that can be as extreme as they are hard to predict. Electricity behaves as a flow commodity, with electricity storage not yet financially or technologically feasible at levels large enough to impact supply behavior. The amount of electricity produced will often depend on hard to predict factors like the weather and results in high disparities between expected and observed prices. There are two types of energy markets, real-time energy markets and day-ahead energy markets. The advantage for day-ahead energy markets is that companies can bid, buy, or sell wholesale electricity 24 hours before the operating day. Companies use these day-ahead markets to avoid some of the market volatility present in real-time markets and reduce their risk. As a result, the modeling techniques for the day-ahead market that companies can use to predict the prices play a large part in reducing the uncertainty involved with day-ahead contracts.

As per a conversation with Green (personal communication, 23 May 2022) helped explain how in Germany, energy is bought and sold on the EPEX SPOT day-ahead market, on which buyers and sellers place bids for physical power volumes to be delivered the next day. Bids are placed by noon the day prior to delivery and are financially binding. The buyer and seller would then use the traded volumes for immediate delivery to their customers. Day-ahead markets differ from real-time energy markets which allow for purchases of energy an hour prior to delivery and expose energy suppliers and purchasers to higher volatility. The resulting outcomes from the day-ahead auction produce the hourly spot prices. These hourly spot prices carry the known hourly price profiles, and the resulting profiles will be used in this paper for modeling. These models, the results of this paper, are used in the Hourly Price Forward Curve (HPFC) for pricing other types of contracts.

The HPFC is constituted of hourly profiles and seasonal profiles calibrated together based on recent energy prices, a technique suggested by Green (2014) and Crispin and Jacobsson (2007). Hourly profiles account for patterns and factors that can affect a given hour in a 24 hour day period. Seasonal profiles account for patterns and factors that can affect a given month in a year. The outputs of both profiles are weights for a given period that are used in the calibration procedure to calculate expected prices. The objective of this thesis is to model the hourly profiles in the German day-ahead market, with a special emphasis on the specific impact renewables have on the HPFC. For a more comprehensive discussion of how profiles are taken together in the calibration procedure, see Green: ‘*A Power Market Forward Curve with Hydrology Dependence - An Approach based on Artificial Neural Networks*’.

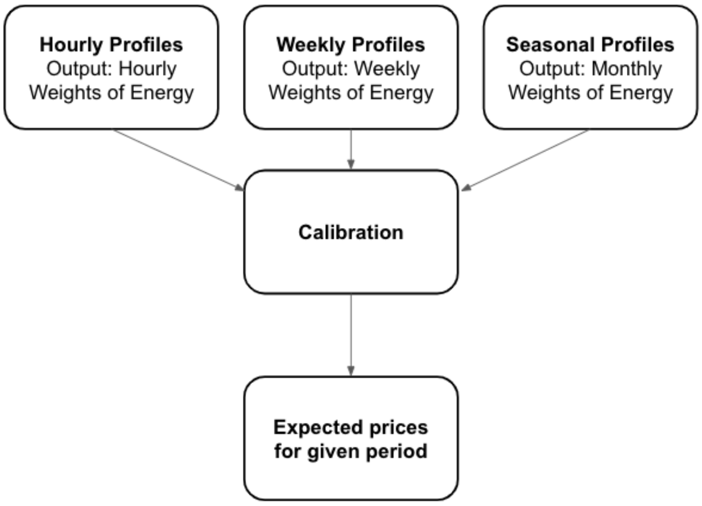

Figure 1.1: Flowchart of modeling process

In a general setting, the hourly profiles will be responsive to daily patterns of human behavior on the demand side and to energy generation patterns on the supply side. Demand, or consumption of energy, differs mainly due to intra-daily, intra-weekly, and seasonal consumption variation, resulting in known patterns. Accounting for patterns on the demand side will be discussed later, but it is important to introduce here the impact of renewables on how the supply side behaves. In the German market, energy is supplied with regard to a merit order, which is a ranking of energy supplies based on marginal cost (Ketterer, 2014). This system sees production taken first from sources which have the lowest marginal cost to produce additional units, placing renewables like photovoltaics (solar), wind, and hydro first in the merit order due to their low (or in some cases negative) marginal cost of production (Beolet, de Jong, & Enev, 2014). In any given market, this will lead to solar, wind, and hydro energy being sold off before other sources that have higher marginal cost. Additionally, the volatility of wind and solar energy production has a disproportionate impact on hourly profiles as a result of wind and solar production being controlled in the short term only by environmental conditions (Ketterer, 2014). This is in contrast to fossil fuel production which can change production in the same direction as price. While some sources of energy can move production levels with price to mitigate volatility, wind and solar produce energy irrespective of demand and price and can exacerbate volatility in the market. This effect has led to wind and solar being dubbed variable renewable energy (VRE) sources, a term to differentiate wind and solar as a result of their volatile characteristics (Rintamäki, Siddiqui, and Salo, 2017). VRE importance for modeling will only increase as wind and solar photovoltaic power generation have the largest projected growth rates among renewable energy systems (Buchholz, 2018).

Outside of a general market, there are factors to the German energy market that make it a distinct setting to model. The share of renewables in Germany accounts for a large percentage, 19.7%, of gross energy production (Umweltbundesamt, 2021; Appunn, Haas & Wettengel, 2021). Importantly, the share of renewables is not allocated equally across energy consumption sources, with renewables accounting for 41% of electricity, 16.5% of heating, and 6.8% of transportation energy consumption in 2021 (Umweltbundesamt, 2021). Additionally, Germany is uncommon in its share of electricity generation from wind and solar, which accounted for 27% and 10.5% of electricity generation respectively in 2020 (Kost et al. 2020; Appunn, Haas & Wettengel, 2021). It is unusual to have a large market where wind and solar production account for this large of a proportion of energy production (DeLuca, 2018; Rintamäki, Siddiqui, & Salo, 2017). This highlights the German energy market's unique characteristics, where VREs are much more volatile in terms of absolute production in comparison to other renewables, like hydropower, or fossil fuels (Xu, Gao, Qian, & Li, 2022). As this relates to hourly profiles, it should be expected that hourly weights will be affected by the absolute production of renewable energy, as well as the expected use for other non-electric energy requirements.

Modeling the hourly profiles will then involve models being able to account for many questions: What are absolute energy production levels expected to be?; What percentage of energy production is required for electricity as opposed to heating?; What is the expected level of low marginal cost renewable energy production?; How will consumption be impacted by perceptions of energy cost?; Are hourly shapes differentiated during a shock to energy demand such as the COVID-19 lockdowns?; etc. Extending from these questions, the hypothesis for this paper is that for the period studied, the share of VRE generation, whether there was a national lockdown for the COVID pandemic, and oil price will all play a complex and substantial part in determining energy price levels. The impact of oil price, COVID lockdowns, and the share of renewables will be the focus of analysis and reported in *Results*. The paper will attempt to answer the question whether VRE related generation alone, along with clock inputs, and further exogenous inputs can significantly capture volatility in the hourly shape profiles of the German EPEX Spot market to model hourly price forward curves.

This paper will attempt to answer this by first providing an analysis of data from EPEX SPOT on energy prices and energy generation, then comparing models of hourly profiles from linear regression, lasso regression with polynomial transformations and variable interactions, gradient boosted regression with decision trees as weak learners, and regression using a feed-forward neural network. The comparative success in modeling daily profiles, as well as insights from the models will be discussed to attempt to answer the research question.

Section 2: Literature Review

Renewable Energy’s feasibility in the energy market

To understand the importance of modeling the impact of renewable energy, first we must assess VRE’s feasibility for electricity in the energy market compared to conventional energy production sources. For the purpose of this paper, Variable renewable energy (VRE) will only include solar, also known as photovoltaic (PV) energy, wind onshore and wind offshore generation. Such sources like gas, hard coal, and coal lignite historically have been viewed as cheaper and more reliable sources of energy, especially if you do not consider the negative externality costs associated with fossil fuel production. The Levelized Cost of Electricity (LCOE), the measure of sum price over the lifetime of energy production, is seen in equation 2.1.

$$LCOE_{type(€cent/kWh)} = \frac{Total\ Lifetime\ Cost_{€cent}}{Total\ Lifetime\ Output_{kWh}} \tag{2.1}$$

The differentiation in technological learning rates that renewable energy sources have compared to conventional sources can be seen in Figure 2.1 from the Fraunhofer Institute (Kost, Shammugam, Fluri, Peper, Jülch, Nyugen, & Schlegl, 2018). Four years ago, the LCOE of renewable energy generation was comparable to or more expensive than conventional sources like coal lignite and natural gas as seen in figure 2.1. In 2021, all non-battery solar and wind generation now have a lower LCOE than most conventional sources as seen in figure 2.2. Renewable energy is currently cheaper in terms of LCOE, which lowers reliance on unstable fossil fuel production in foreign nations and produces less negative externalities associated with fossil fuel production.

2018 LCOE in Germany (renewables circled)

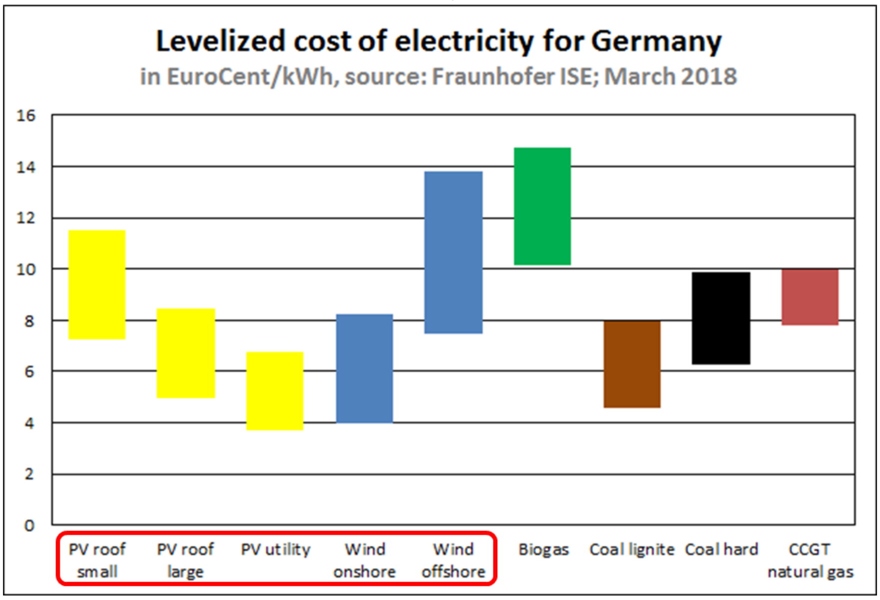

Figure 2.1: Levelized cost of electricity change in Germany 2018 (Kost et al. 2018)

2021 LCOE in Germany (renewables circled)

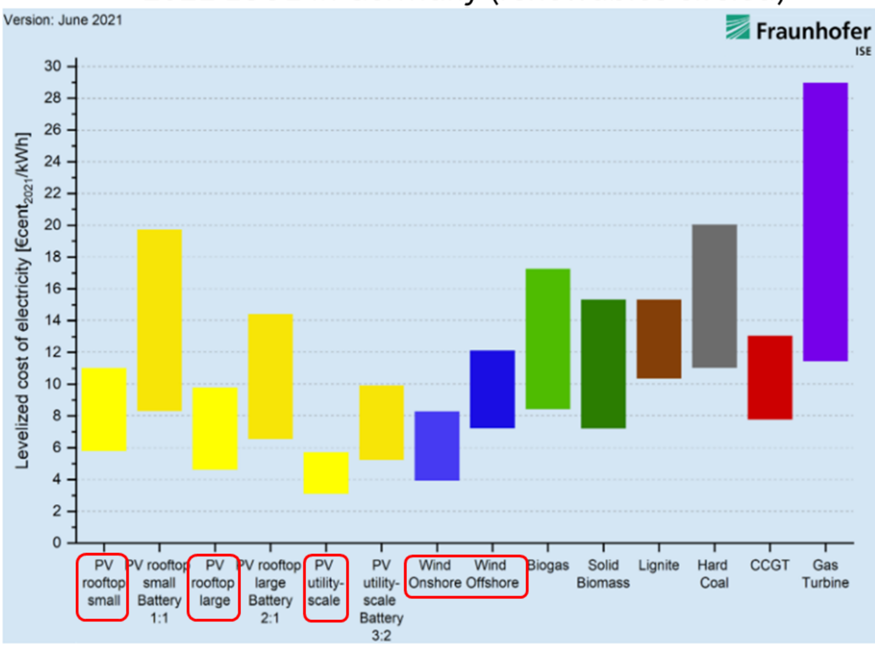

Figure 2.2: Levelized cost of electricity in Germany 2021 (Kost, Shammugam, Fluri, Peper, Memar, & Schlegl, 2021)

According to the Kost et al. (2021) report on the LCOE of renewable energy technology, renewable energy production has a lower levelized cost of energy than conventional sources, namely biogas, biomass, coal lignite, and hard coal. Photovoltaic energy, depending on it's scale of production in Germany has a LCOE between 3.12 and 11.01 €cent/kWh, Wind onshore 3.94 and 8.29 €cent/kWh, and Wind offshore between 7.23 €cent/kWh and 12.13 €cent/kWh (Kost et al. 2021). Overall Utility Scale Photovoltaic has the lowest cost of any electricity source as seen in Figure 2.2, and solar will only become cheaper with an assumed technology specific learning rate of 15% for photovoltaic energy. (Kost et al. 2021). On figure 2.2, it can be seen that even the cheapest form of fossil fuel electricity production coming from Combined Cycle Gas Turbine (CCGT), now has become relatively more expensive in Germany than almost all solar energy production and wind energy (Kost et al. 2021). For the purposes of building accurate predictive models, these results suggest it is critical to understand the effects that renewable energy generation has on the shape of the hourly spot price curve.

Appropriate Time Window of Study

For this paper we will use the last 4 years of data ranging from 31 January 2018 to 31 January 2022. The first reason for such a constraint is that the proportion of renewable energy in the German energy market has been more substantial within the last few years, which has a pronounced impact on the day-ahead spot price when compared to previous daily shape curves (Beolet, de Jong, & Enev, 2014). The second reason for only including data from the last four years is due to the decreasing renewable energy prices driven by a high technology specific learning rate within renewables (Duffy, 2020). For example, wind onshore production in Germany saw a 33% decline in LCOE in 2020 (Duffy, 2020). Internationally, there has been a trend toward larger, taller wind turbines with higher capacity factors for both onshore and offshore wind production (Duffy, 2020). The recent reduction in the price of renewable energy is also a result of economies of scale for renewable energy infrastructure (Holm & McIntosh, 2008), German government incentive programs for renewable energy (Matschoss, 2018), and increasing prices of CO2 certificates in Germany (Kost et al., 2021).

VREs are noted to have a direct impact on German energy market prices. De Jong (2013) showed how photovoltaic energy largely influenced daytime energy pricing as a 10% increase in electricity production from renewable resources during daytime hours leads to a 6.6% reduction in electricity prices. The effect of photovoltaic production has likely increased in scale of effect as in 2013 wind and solar generation in Germany only accounted for around 15% of the electricity market (Matschoss, 2018) compared to 41% of electricity in 2021 (Umweltbundesamt, 2021). Therefore the effect of VREs in the current time window should have an even more pronounced effect.

Renewable Energy and Exogenous Variable Influence on Day-Ahead Energy Market

One key paper, *Effects of Renewables in the Stylized Facts of Electricity Prices*, by Ballester and Furió (2015), provides a foundation on the negative relationship between renewable generation share and the day-ahead market marginal prices in Spain, and that the generation share volatility is transferred to price volatility. Additionally, Ballester and Furió

(2015) gave evidence of the effect of renewable generation to drive down day-ahead prices, even with the higher cost of renewables in 2015. Their research provides good evidence that renewable energy generation has a negative correlation on price and that the stochastic generation of renewable energy brings forth the price volatility in the hourly price profiles which this paper attempts to capture with the models.

Further evidence of renewable's effects on the intra-daily shapes on day-ahead spot prices specifically within the German market can be seen in the KYOS Analysis Report named *Improved Hourly Shaping using Renewable Production Information* (Beolet, de Jong, & Enev, 2014). The report depicts historical spot prices in Germany from 2001 to 2013, figure 2.3, in order to show the effects of hourly spot price patterns, especially concerning how solar is related to a steeper price drop in 2013 compared to previous years (Beolet, de Jong, & Enev, 2014). The paper also reports a 1% increase of wind generation is related to a 1-5% decrease in power price in Germany, while power prices are also shown to be dependent on solar production during daylight hours (Beolet, de Jong, & Enev, 2014). Additionally, the KYOS report predicts that later years saw the renewable component have a larger impact for longer-term curves (Beolet, de Jong, & Enev, 2014). Therefore, the data for this paper from 2018-2022 is expected to yield an even larger impact from renewable energy on the intra-daily shapes.

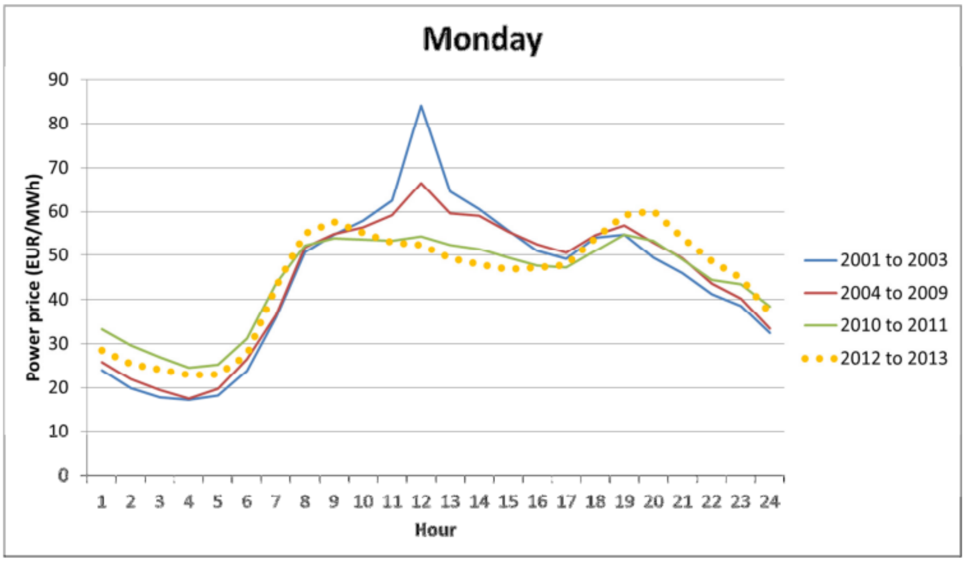

Figure 2.3: Historical spot prices on average monday in Germany (Beolet, de Jong, & Enev, 2014)

Many variables were considered for inclusion within tested models to account for macroeconomic factors into the spot prices. Previous research by Jones (1996) provided evidence that macroeconomic information is encapsulated by oil prices, specifically when oil price shocks occur. Jones discussed how oil prices, specifically oil shocks, are indicators of disruptive macroeconomic trend changes and how oil prices contain asymmetry of macroeconomic responses for recessions (Jones, 1996). Oil also now competes with electricity for transportation and heating, so often when oil prices increase or decrease, energy prices follow suit. Therefore it was considered to be the appropriate inclusion as a covariate into the model.

Autoregressive Time Series Models

In previous modeling of spot price and price volatility, Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) algorithms have been popular due to the general characteristics observed in the energy market (Frömmel, Han, & Kratochvil, 2014; Rintamäki, Siddiqui, & Salo, 2017; Ketterer, 2014). An autoregressive approach is a natural fit for predicting price over time. Although research using these methods have highlighted potential shortcomings (Frömmel, Han, & Kratochvil, 2014; Rintamäki, Siddiqui, & Salo, 2017; Ketterer, 2014).

Autoregressive based research in the form of a seasonally adjusted autoregressive model using variable renewable energy in the German market provided evidence for the pronounced impact of wind and solar on spot price volatility (Rintamäki, Siddiqui, and Salo, 2017). However, Rintamäki, Siddiqui, and Salo's (2017) use of a distributed lag model emphasized interpretable coefficient impacts, rather than predictive accuracy. Their seasonally adjusted autoregressive model was able to draw conclusions regarding the relationship of VRE's on price, but different approaches would be needed to capture the energy market's complexity involving high peaks.

Using a more flexible approach, Frömmel, Han, and Kratochvil (2014) drew important conclusions regarding modeling energy price volatility in the German market using a GARCH model. Core to Frömmel, Han, and Kratochvil's (2014) methodology were specific stylized facts; namely large price volatility and long persistence of the volatility. The authors also noted high mean reversion of price, multiple seasonality and a stationary price level, the last point noteworthy because it was not observed in the period studied for this paper (Frömmel, Han, & Kratochvil, 2014). In their research, a GARCH approach was pursued as opposed to more standard autoregressive algorithms because of the volatility characteristics of energy price, with the authors suggesting that the characteristics of traditional autoregressive algorithms were ill suited towards capturing the spike characteristics in mid to long term projections. To circumvent this, a realized GARCH model with a skewed t-distribution was developed to better account for spike characteristics, but was not able to necessarily improve prediction on out-of-sample forecasts (Frömmel, Han, & Kratochvil, 2014).

Research conducted around the same time as Frömmel, Han, and Kratochvil (2014), Ketterer (2014) drew similar conclusions regarding the shortcomings of GARCH with regard to volatility. Ketterer (2014) analyzed price levels and volatility in the German market, concluding that wind production has a negative effect on prices but increases price volatility. This research also confirmed that the merit order effect, which can be thought of as the downward effect on prices from VRE, is stronger during the day (Ketterer, 2014). Ketterer's intent was to grow the understanding of VRE's impact on price and volatility and the author notes that predictive modeling using GARCHs would need extensive exogenous variables to enhance the predictability of volatile episodes.

Complimentary research done on observations of negative energy prices in the German market by Genoese, Genoese, and Wietschel in 2010 highlight both how extreme the periods of volatility can be for the energy market and how VREs are the most critical factors to understanding those periods. They provide evidence that wind generation was found to be the most important factor related to the observation of negative prices, of which there were

86 negative spot prices between September 2008 and December 2009. Similar research by Gonzalez-Aparicio and Zucker (2015) in the Spanish market confirmed wind volatility as critical to accurate forecasting. In the period studied for this paper, 741 observations of hours with negative prices were observed over a period of 4 years for the German market, with certain periods exhibiting higher rates of negative prices than Genoese, Genoese, and Wietschel (2010).

An alternative approach to autoregressive models was concurrently developed based on modeling time spot price profiles over a given period and combining them in a calibration procedure which tailors these profile predictions to prevailing forward prices (Green, 2014; Crispin & Jacobsson, 2007). The profile calibration design with hydrological balance as a key input was shown to account for price spread among low and high hydrological balances over 10 years in the Nordic market by Green (2014). Subsequent research into VREs and further exogenous variables in order to apply the profile calibration modeling procedure to the more volatile German EPEX Spot market is the basis of this paper.

Outlook on the Future of the Energy Market

As will be discussed in the *Data* section, it is important to note the importance of model validity in the future. Beolet, de Jong, and Enev (2014) outlined the impact renewables have on prices, while renewable energy has nearly tripled in absolute production (Matschoss, 2018). By 2030, new photovoltaic panel installation and wind installations could be cheaper than running existing conventional power plants (Kost et al. 2021). According to Hansen (2018), there is evidence that Germany is undergoing a full energy sector transition to become 100% renewable energy base by 2050. The goal of 100% renewable generation is not limited to electricity, but also includes heating, industry, and transport (Hansen, 2018).

The proposed models for this research will not limit our focus on conventional sources of electricity which may result in more accurate predictions, but may become irrelevant in the future of the market. Conventional sources are adjusted when demand and price require additional energy and could be poor indicators for forecasting price in the future. Including all conventional sources of energy would create more noise, adapt less to days which have a highest share of renewable energy production and, as Hansen (2018) suggests, would create a largely unimpactful model for the future of the EPEX SPOT day-ahead market.

Section 3: Data

Outline of SMARD.de and Other Data Sources

The data used for this thesis is for the period of 00:00 January 31 2018 to 23:00 January 31 2022. The main source for the data is SMARD.de, an electricity market information platform run by the Bundesnetzagentur in Germany (SMARD, 2022). SMARD is an abbreviation of a German term referring to electricity market data. From SMARD, market data of hourly spot prices was pulled, along with information on hourly generation for solar, wind, nuclear, biomass, hydroelectric, hard coal, lignite, hydro-pumped storage, and other fossil fuels. Consumption data was also pulled from SMARD, consisting of total grid load, and hydro pumped storage consumption data. Information regarding daily crude oil prices in the EU was pulled from the ALFRED archival economic data website and information on Germany lockdown requirements during the COVID pandemic was pulled from a variety of sources (Miller, 2020; Seythal & Carrel, 2020; DW News, 2020; Sky News, 2020; DW News, 2021) as there is not yet in place an official German lockdown database. Information on German national holidays was taken from the Workalendar package for Python (Bord, 2022).

It is important to note that the actual historical energy generation data was used for this study, as opposed to the forecasted historical generation data, which is also available on SMARD. Models created for this study are intended to pick up on true relationships that affect hourly profiles, rather than artificial relationships possibly developed as a result of forecasting models. Actual energy generation data is self-evidently only available ex-post, and can not provide a true idea of the performance of these models in forecasting. However, the disparity between forecasted and actual energy generation for solar and wind are being refined and can be expected to receive more research in the future to reduce this disparity (Gonzalez-Aparicio & Zucker, 2015). This subject will be explored further in the *Limitation* section of *Results*.

In this paper the hourly profiles will attempt to capture yearly, monthly, and daily patterns into a single output vector which can be used with predictions of the average spot price per hour for a given month. The models attempt to account for time patterns through the inclusion of the weekly, seasonal, and hourly clocks. In theory, this allows the target vector to be detrended and resilient to shocks, such as the jump in energy prices in Germany in the winter of 2022.

Feature Engineering and Variable Selection

Spot prices are registered on the German market in Euros per Megawatt hour(€/MWh). Unlike previous papers using different time periods which observed a high degree of mean reversion, the spot prices between 2018 and 2022 show an upwards trend in figure 3.1. Figure 3.1 displays the trending price, beginning with a slight increase in price beginning in October 2020, then a rapid increase in price from September 2021.

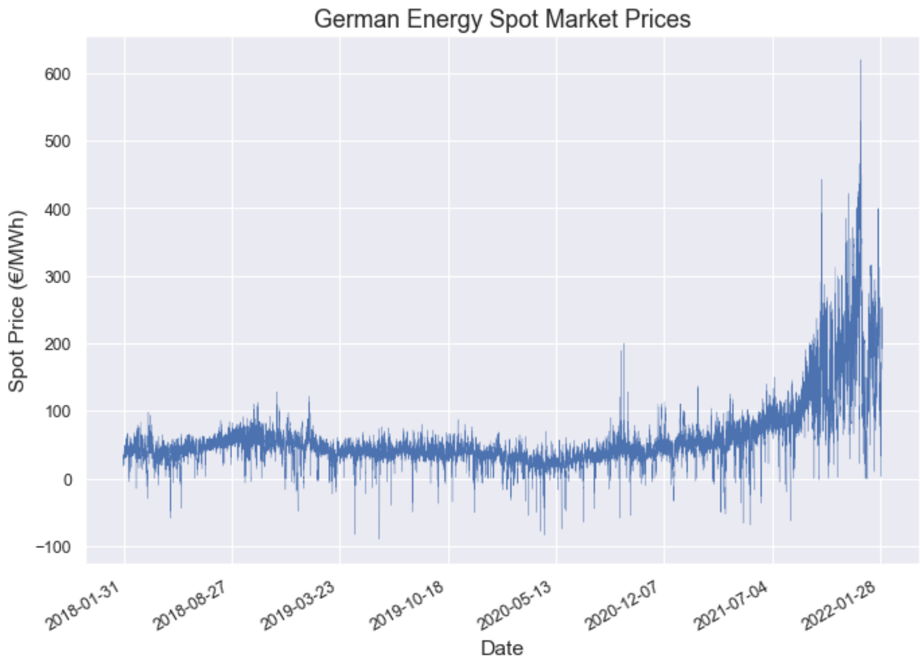

Figure 3.1: Observed hourly spot prices between 2018 and 2022

To remove longer-term trends, which should be accounted for in the complementary models to hourly profiles, the Hour-to-Month ratio or H2M ratio, equation 3.1, is used as a target vector.

$$H2M_h = \frac{P_h}{\overline{P_{h,m}}} \tag{3.1}$$

Many explanatory variables were taken from SMARD.de, including 13 different types of energy generation and three energy grid load variables, most importantly residual load. Residual load can be understood as the energy capacity that is left, after renewable sources have cleared, that can be used for conventional power plants to supply electricity. Residual load increases as the capacity of VRE grows, VRE power output begins to affect the load balance of the power system and was first defined by Bofinger, Saint-Drenan, von Oehsen, Gerhardt, Sterner, & Rohrig (2009) for the German market. Residual load encompasses conventional energy sources. Consequently, inclusion of specific conventional generation in our data could create covariance with residual load, while taking away focus from the primary focus on renewable energy. Therefore residual load will only be the only reference to conventional energy generation, and will be included in models as a reference for the proportion of renewable energy. This process will be described in the feature engineering portion of the *Data* section.

New variables were created during a feature engineering phase which were hypothesized to have a significant relationship with the target vector. Most features had little relationship with the dependent H2M ratio, as was the case for the solar-to-wind ratio. The most significant feature created was the sum of VRE generation, wind onshore, Won_h; offshore, $Woff_h$; and photovoltaic energy, PV_h, over the residual load, RL_h, as seen below in equation 3.2. This ratio was suggested by de Jong, van Dijken, & Enev (2013) to be relevant to the behavior of spot price. This ratio will be referred to as the 'renewable ratio', RR_h, throughout the paper.

$$RR_h = \frac{\sum_{i=1}^{N} \left(Won_h + Woff_h + PV_h \right)}{RL_h} \tag{3.2}$$

Another target vector for modeling was initially considered for based target vectors used in previous research, a target vector with daily weight to sum to unity for the day (Green, 2014). Each hour's spot price would be divided by the sum of 24 hours for that day as seen in equation 3.3 represented by $SD_{h,d}$. This would be applied to all hours in all 1462 days during the 4 years of data. The issue identified with this target vector was that it could not easily interpret the effects of the days of the week, and required a second network to compensate for that shortcoming. Further into research, it was discovered that the shift to using the H2M ratio, as seen before in equation 3.1, kept more variation in the distribution of the target vector, and provided more accurate results.

$$SD_{h,d} = \frac{P_h}{\sum_{i=1}^{d} \left(P_{h,d} \right)} \tag{3.3}$$

This modeling environment was felt to not include enough variables exogenous to the electricity market which were identified by Ketterer (2014) as critical for encapsulating spike pricing. New variables were researched, and two variables were implemented to account for these conditions.

A significant portion of the data studied was during the COVID-19 lockdowns. Overall energy demand decreased during these periods (Abu-Rayash, 2020). Therefore, there is a need to specify the dates in which lockdowns are incorporated into the model. For this, a binary variable known as 'COVID' was created to indicate the time of the two national lockdowns. First, the initial shutdown of schools and large events when COVID-19 from 2020-03-10 to 2020-03-16 (Miller, 2020),between 2020-03-16 to 2020-05-10 for the first national lockdown (Seythal & Carrel, 2020; DW News, 2020), and between 2020-11-02 to 2021-03-01 for the second national lockdown (Sky News, 2020; DW News, 2021).

Figure 3.2 displays the distribution of the spot prices during the COVID-19 lockdown in Germany in orange, and no lockdown periods in blue. The dotted lines represent the average price for COVID and non-COVID lockdown periods. The resulting average spot price during the pandemic is 36.6, while it is 57.9 when there was no lockdown. Additionally, the distribution of the prices during the COVID-19 lockdowns appears different from the distribution of normal periods. Lockdown periods exhibit characteristics of a bimodal distribution, while the non-lockdown distribution has a very long upper tail, which extends past the limits of this plot to the maximum price of 620 and pushes the mean of the non-lockdown period much higher. The difference in distribution of the COVID variable could be due to a lack of data points or because there were 2 main lockdown periods, where prices were rather consistent.

The differences in distribution suggests the COVID binary variable could have a significant effect on the dependent spot price. For this study it is important to detrend the data, to show the effect that energy generation has on the price, and apply the H2M ratio as discussed previously. In figure 3.3 the plot compares the lockdown distribution compared for the H2M ratio in both periods. In figure 3.3, COVID-19's effect on price is almost negated when you compare the H2M averages. During COVID lockdown the average hour over monthly ratio is 0.995, while normal periods have an average of 1.005. It may be that the effects of COVID-19 lockdowns itself could have a lagged effect on energy prices, and therefore not all information is captured within the nature of a binary variable. Therefore it may be expected that COVID-19 variable will have a direct impact on the target vector.

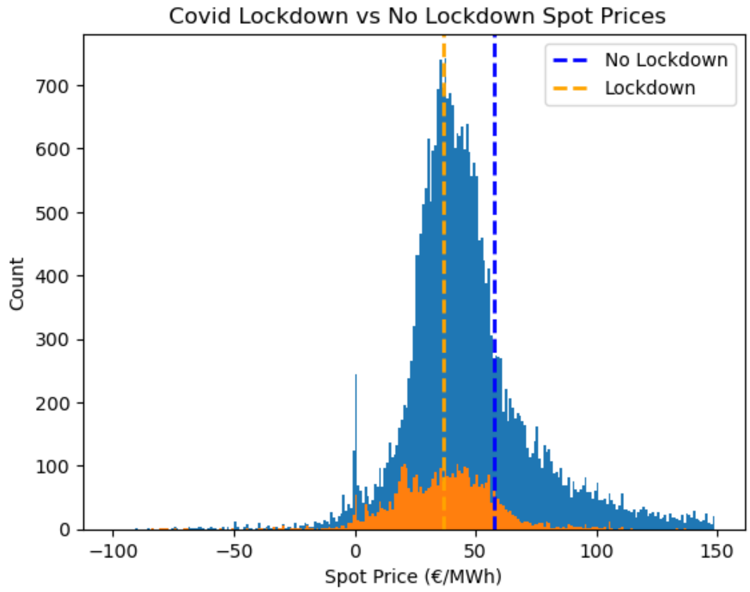

Figure 3.2: Distribution of price during COVID Lockdown and without COVID Lockdown

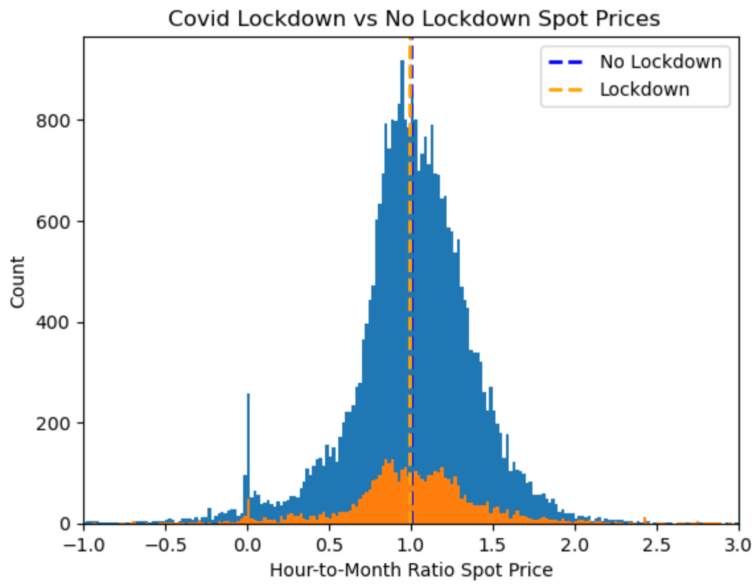

Figure 3.3: Distribution similar to figure 3.2, with H2M ratio instead of spot price

The oil price variable used in modeling represents crude oil prices within the EU, represented in US dollars per barrel (US Federal Reserve, 2022). The data for this variable contains daily prices, and therefore remains for all 24 hours in each day modeled. The data contained NA values, as prices during holidays and weekends were not accounted for. The previous day's prices were used in place of these NA values. Figure 3.4 shows oil prices along with German energy spot prices, both standardized. A positive correlation between the two prices for the beginning of 2021 to the end of the data period suggests that oil price could be increasing in importance for prediction in the future.

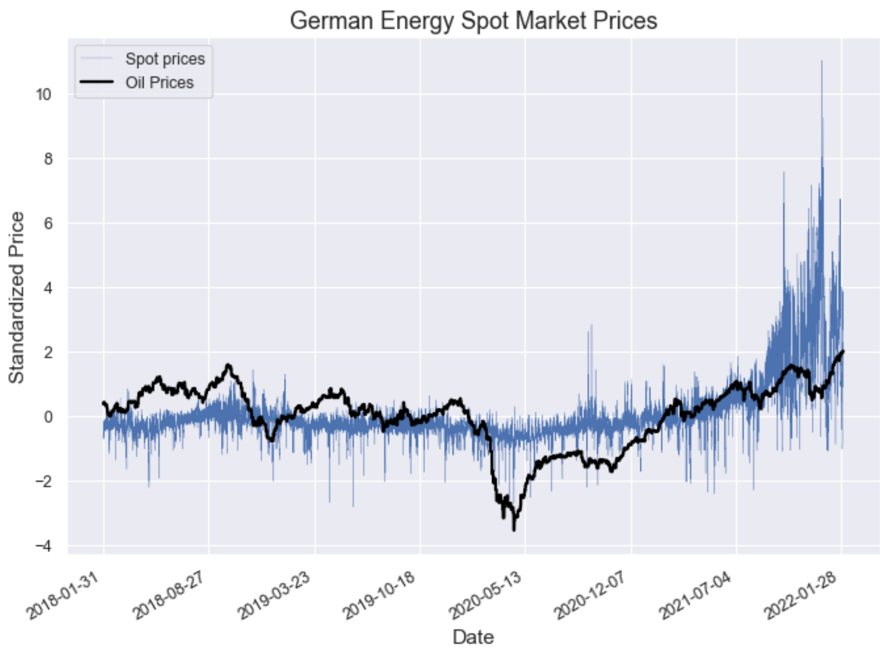

Figure 3.4: Standardized oil price with standardized spot price

At the beginning of our paper we sought to do polynomial transformations with all generation and consumption variables from SMARD.de, along with the features engineered as previously noted. By including all variables along with their polynomial transformations into a Lasso regression, the L1 norm removed insignificant variables as a tool for variable selection, as recorded in other studies (Fonti, 2017). This variable selection technique was used to analyze the possible significance of engineered features, and suggested that the renewable ratio, VRE over residual load, would be important for modeling.

Figure 3.5 gives box plots of the variables that will be used for modeling. Here total grid load is also displayed as a representation of energy demand, to represent the fluctuation in the total demand for energy. As displayed in the distribution it is to be noted that it is quite consistent, and will mainly vary by time of day, day of the week, and time of the year. Other variables are included on the plot showing much more drastic variability, and a good amount of outliers. The amount of outliers within the target shape vector H2M ratio is quite extensive, and should be accounted for in the methodology of the models. Additionally the renewable ratio has a very large amount of outliers as well, signifying days where renewable energy production is drastically greater than average. These days will be discussed later in the discussion section.

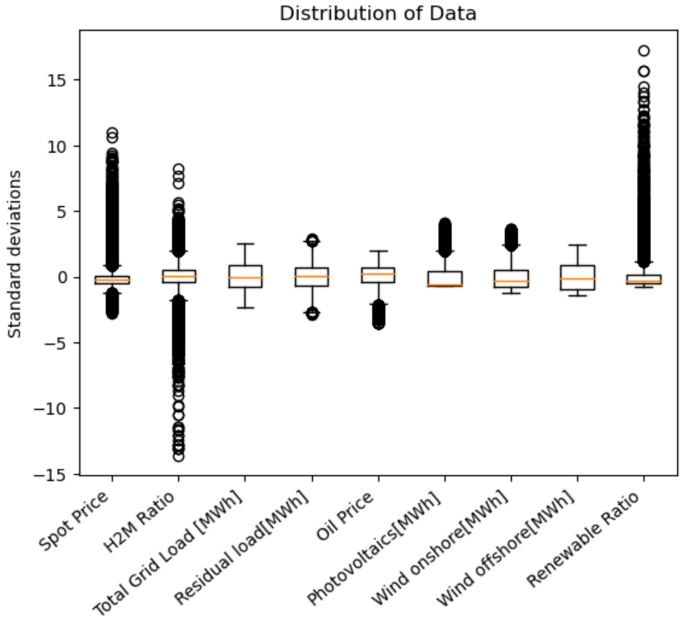

Figure 3.5: Boxplots of input variables and total grid load

Intra-daily and Intra-weekly Shapes

Stylized facts presented by Green (2014) are reevaluated for the German market for the period between 2018 and 2022. Green (2014) showed that day types exhibit characteristic differences that change hourly profiles . Figure 3.7 gives the average weights for each day for the German market between 2018 and 2022, so that the weights sum to one. There is a clear drop in weights for Saturday and Sundays, in line with Green's (2014) findings for the Nordic market between 2002 and 2011.

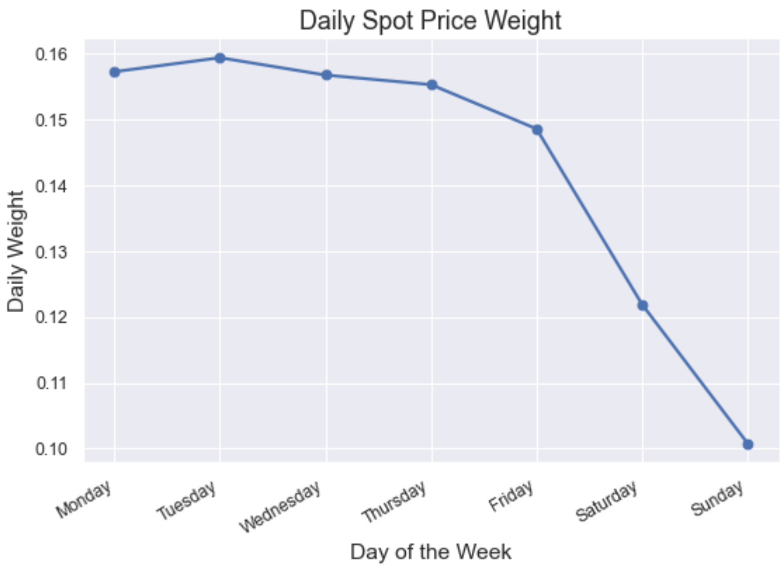

Figure 3.6: Weights of average hourly price for each day of the week

Green (2014) also presents the stylized fact that hourly profiles of energy will behave differently depending on the day of the week. Figure 3.7 is taken from Green (2014) and shows the weights of average energy prices for each hour by daytype, normalized to sum to one, for a period between 2002 and 2011 in the Nordic market. To confirm this stylized fact in the German energy market between 2018 and 2022, the same graph is presented in figure 3.8.

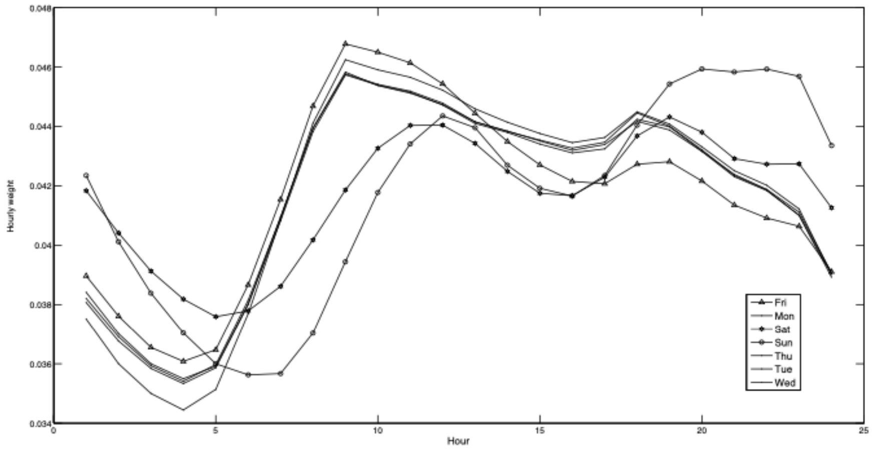

Figure 3.7: Hourly profiles of different days of the week for the Nordic market between 2002 and 2011(Green, 2014)

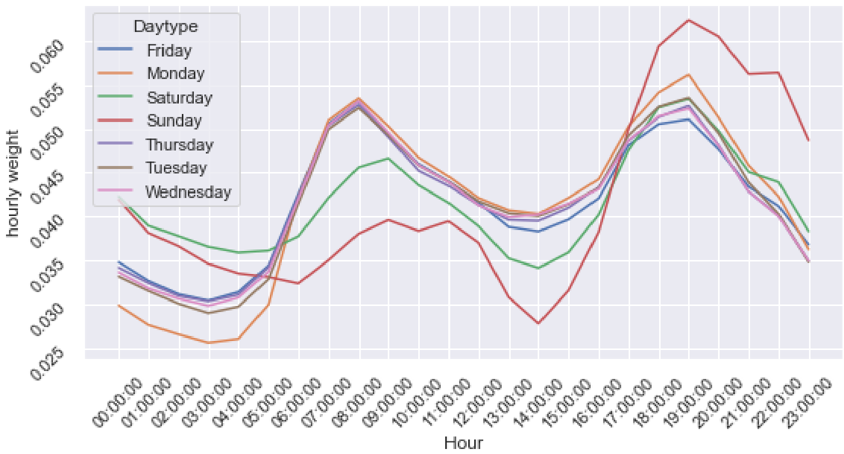

Figure 3.8: Hourly profiles of different days of the week for the German market between 2018 and 2022

Both figures demonstrate that each day of the week has a different level of consumption, which then in turn has an effect on prices. Tuesday-thursday are similar, but all other days of the week behave differently. Day of the week prices have different trends especially during the weekends, with Sunday and Saturday reflecting lower overall prices as seen in figure 3.8. Intra-daily clocks, and intra-weekly clock representations will be included in all models.

There are a few differences worth highlighting between the two figures. First, the German market in the second figure features a more pronounced dip around the afternoon, likely as a result of Germany's higher solar production with respect to the Nordic market. A second key difference is the larger disparity between daytype trends in the German market for 2018 to 2022, as can be seen by the larger gap between lines seen in figure 3.8. This analysis suggests the stylized fact presented by Green (2014) holds, and the greater disparity between daytype weights seen in figure 3.8 suggests that daytypes behave more distinctly in the period of this paper than seen in the Nordic market between 2002 and 2011.

Final Network inputs

The inputs for each model vary slightly given the requirements of each algorithm, but largely attempt to account for the same features, which will be described here. A total of 35084 hours were used for modeling, with 80% being allocated for training. Table 3.1 provides the first 5 rows of the dataframe used to train the linear model. Table 3.2 gives description statistics for the dataframe used to train the linear model.

Table 3.1: Heading of inputs included in each algorithm

Unnamed: 0	Yearly Clock (sin)	Clock (daily weight)	Photovoltaics[MWh]	Wind onshore[MWh]	Wind offshore[MWh]	PV + W / Residual Load	COVID	Oil Price
0	0.0	0.033669	0	3030	987	0.411831	0.0	67.78
1	0.0	0.031862	0	3351	1009	0.497490	0.0	67.78
2	0.0	0.030679	0	3736	1055	0.594048	0.0	67.78
3	0.0	0.029853	0	4370	1064	0.726762	0.0	67.78
5	0.0	0.033733	0	5377	1084	0.864348	0.0	67.78

Table 3.2: Summary statistics for each input in the training set

	Yearly Clock (sin)	Clock (daily weight)	Photovoltaics[MWh]	Wind onshore[MWh]	Wind offshore[MWh]	PV + W / Residual Load	COVID	Oil Price
count	28066.000000	28066.000000	28066.000000	28066.000000	28066.000000	28066.000000	28066.000000	28066.000000
mean	0.633752	0.041658	1258.002280	2735.826730	681.777132	0.690429	0.136464	62.371788
std	0.316695	0.007687	1924.393817	2171.978183	466.667000	0.882507	0.343287	14.918172
min	0.000000	0.025666	0.000000	25.000000	0.000000	0.005721	0.000000	9.120000
25%	0.500000	0.035912	0.000000	1065.000000	252.000000	0.212569	0.000000	55.640000
50%	0.707107	0.041613	23.000000	2081.000000	640.500000	0.428267	0.000000	65.190000
75%	0.965926	0.048101	2040.000000	3860.000000	1069.000000	0.823807	0.000000	72.540000
max	1.000000	0.062358	9068.000000	10558.000000	1804.000000	14.441676	1.000000	92.350000

The yearly clock is an input that identifies the time of the year for the model. Except for the neural network, this is in the form of a sine function sampled at equal intervals 12 times between 0 and pi. This transforms the categorical variable of month into a continuous variable in which winter months are closer than summer months, etc. The second input, daily clock, informs the model of the given time of the day. Different methods are used to build the daily clock for each algorithm and will be discussed in the methodology section, but figure 3.4 presents a graph of two different daily clocks explored, a sin function similar in design to the yearly clock previously explained and the average daily weight for a given hour of a certain daytype (in this case a Wednesday).

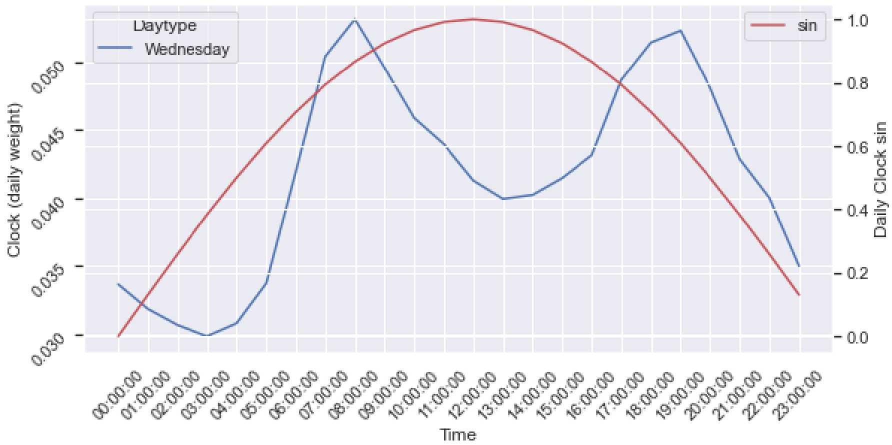

Figure 3.9: Hourly clocks from a sine function and average daily weights

The next three inputs are taken directly from SMARD.de as previously discussed, photovoltaic, onshore wind, and offshore wind production, all in megawatt hours (MWh). The sixth input is the renewable ratio as previously noted. The seventh input is a binary class variable that indicates whether Germany was under a federal government mandated lockdown during that specified hour. The last variable is the oil prices.

Renewable Energy and H2M ratio

To explore the central object of this study, the relationship between hourly energy price and renewables, a pair of scatterplots are presented in figure 3.10. The graphs show solar and wind offshore generation on the y axis, while the x axis is the observed spot price over the average spot price for the month.

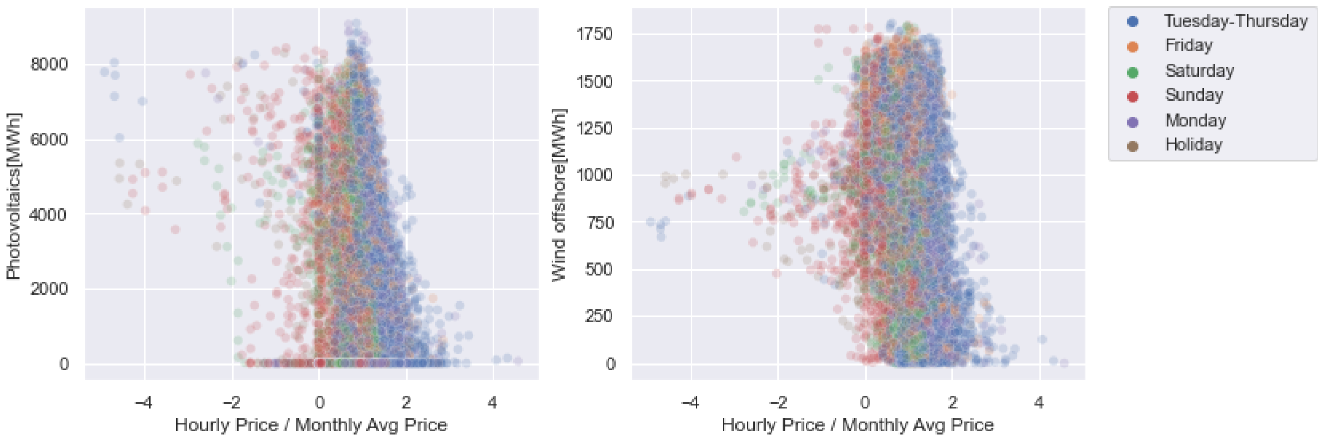

Figure 3.10: Scatterplots of VRE variables and detrended spot price

From the graph it is not immediately clear whether there exists a relationship between the renewable generation sources and the deviation of spot price from the monthly average. Arguably, the only clear pattern is a grouping around 1 on the x axis, near the monthly average spot price, in line with the stylized fact that the spot price features a high degree of mean reversion, suggested by Frömmel, Han, and Kratochvil (2014).

A more interesting relationship is shown between the renewable ratio and H2M ratio, shown in figure 3.11**.** A complex and distinct relationship can be seen between a drop in the share of renewables and a rise in the hourly spot price. This relationship should be expected, but less intuitive is the large heteroskedasticity as the renewable ratio increases. Figure 3.11 gives an indication that complex relationships exist between the hourly energy price and the share of renewables, that will require a dynamic algorithm to adequately capture.

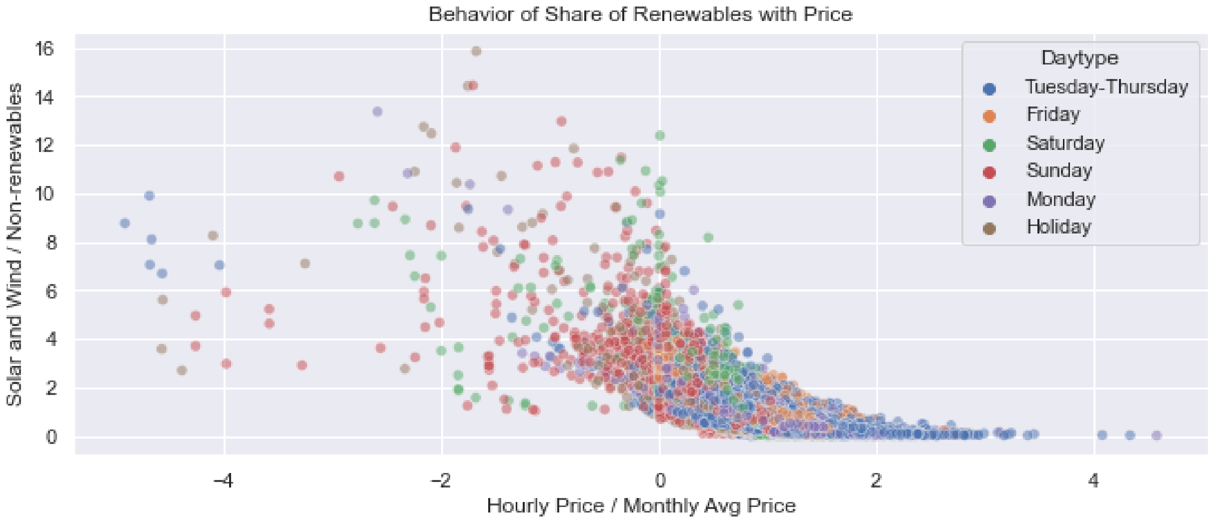

Figure 3.11: Scatterplot of the share of renewables and detrended spot price

Section 4: Methodology of 4 Algorithms

After defining the final input variables and the target vector, we investigated 4 different algorithms to best create a shape curve to use for the day-ahead energy market in Germany. The four algorithms were linear regression, linear regression with lasso penalization and polynomial transformations and interactions, a feed-forward neural network (FNN), and gradient boosted decision trees. Each algorithm will have its own detailed methodology.

Linear Regression

The baseline algorithm chosen to start the modeling was linear regression. Linear regression gives predictions based on explanatory variables with the assumption of linearity, that increases or decreases of inputs produce scalar increases or decreases in the output. Each explanatory variable is thereby associated with a coefficient value which informs the model of how an increase in that variable will affect the output. Linear regression can be expressed as equation 4.1 where the matrix, $\mathbf{X}$, is transposed such that the multiplication with the vector of coefficient values, $\boldsymbol{\beta}$, occurs with their associated regressors and are summed to reach the output (Hastie, Friedman, & Tisbshirani, 2009). These coefficient values are such that they minimize equation 4.2, which is the summed distance between observed and estimated outputs, or residual sum of squares (Hastie, Friedman, & Tisbshirani, 2009). Such estimates of these coefficients are developed on training data, provided the input matrix is nonsingular, by taking the inverse of the inner product of matrix $\mathbf{X}$ and its transpose with $\mathbf{X}^T$ and output of the training observations y, equation 4.3 (Hastie, Friedman, & Tisbshirani, 2009).

$$y_i = X_i^T \beta \tag{4.1}$$

$$RSS(\beta) = \sum_{i=1}^{N} (y_i - X_i^T \beta)^2 \tag{4.2}$$

$$\hat{\beta} = (X^T X)^{-1} X^T y \tag{4.3}$$

Certain manipulations of our data were needed for use in a linear model. In particular, variables relating to informing the model of the time, the hour of the day and the day of the week. Two forms of the daily clock were explored for use with the linear model, both discussed in the section *Data*. One daily clock to inform the model of the time of day was 24 hourly samples from 0 to π of a sine function with a periodicity of 48 hours. The other daily clock, which slightly outperformed the sine clock, was the average weights of hours for given daytypes, shown in equation 4.4. The average weight clock informs the model of time of day by how much weight a price for a given hour for a given daytype has on average in the sample period.

$$x_{h,d} = \frac{\bar{p}_{h,d}}{\sum_{i=1}^{24} p_{h,d}} \tag{4.4}$$

Examples of both these daily clocks are provided in figure 3.9 in the section *Data*. The day of the week, a categorical variable, was given to the linear model as one-hot vectors, where day of the week was converted into seven variables where only the relevant weekday variable is assigned a value of 1 and the rest are left with values of 0.

Lasso with Polynomial Transformations and Interactions

The second algorithm explored was a lasso regularized linear regression which was given the same linear regression variables, as well as degree 2 polynomial transformations of those variables and multiplicative interactions between the original variables. LASSO, or Least Absolute Selection & Shrinkage Operator, is a regularization method which assigns a constraint on the sum of the absolute values of the coefficients (Hastie, Friedman, & Tisbshirani, 2009). Seen in equation 4.5, estimates of β coefficients with Lasso regularization are reached by minimizing the residual sum of squared residuals plus a penalty term which is the product of λ, a hyperparameter, and the sum of the absolute value of coefficients (Hastie, Friedman, & Tisbshirani, 2009).

$$\hat{\beta}^{Lasso} = argmin_{\beta} \left\{ \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij} \beta j)^2 + \lambda \sum_{j=1}^{p} \left| \beta_j \right| \right\} \tag{4.5}$$

The value of λ controls the strength of penalization, where there is no penalization when λ equals zero. For the case of this paper, the optimal λ value to minimize error was chosen via 5 fold cross validation with 10000 iterations, ending up at 7e^-4.

Lasso penalization allows for reduced variance between samples with small increases to model bias and can help with overfitting. Additionally, as the name suggests, lasso can be used as a selection tool of variables, providing insight into relevant predictors (Hastie, Friedman, & Tisbshirani, 2009). These characteristics allow lasso regularized models to be given a wide range of variables and use only the variables that are relevant to prediction. For this reason, the lasso model was given extra transformations of inputs in the form of degree two non-linearities that had better explanatory ability than the linear trend. Additionally, which regressors were selected by the lasso model could provide general insight for modeling the spot price when using the provided variables.

Neural Network

This paper had a lot of its initial research based off of the previous paper by Green (2014). He utilized an Artificial Neural Network, or more notably a feed-forward Neural Network. As previously discussed, he utilized the combination of an hourly network (where the target vector was represented by the price of the hour divided by the sum of all 24 hours price for that day) and a daily network, (which accounted for the noticeable different shape curves that each day of the week). The daily model used a 6-60-24 feed forward model with 24 nodes for the output layer representing hours of the day, and the daily network used a 10-60-7 feed forward network. For the network we initially attempted to replicate a similar version of the model outlined by Green (2014).

When constructing this neural network, initial issues were made as no matter how the model architecture was formed, the models could not account for low variability within the daily target vector, and predictions did not accurately predict the daily sum target vector. One fix was applying a logarithmic transformation to the target vector as the values should have all been between 0 and 1. This was successful at improving the neural network's predictive accuracy, although it required removing a sizable proportion of the data as negative prices caused values to go below 0, or above 1. Therefore, a logarithmic transformation could not be applied, as negative price data constituted meaningful data points which needed to be accounted for in modeling. As a result, the H2M ratio was implemented as the target vector.

For the neural network, unlike previous models, the day of the week, a categorical variable, was not fed into the neural network as a one-hot encoded categorical variable. Instead a sin/cosine function stood inplace to represent the weekly clock. Additionally, the time of the year as in previous models was given a cosine model which helped to represent the time of year, which the neural net could register. This technique was used by Green (2014) and in practice, did indeed register better results than categorical variables used in the alternative models.

Initially grid-search cross-validation was implemented to select the best parameters in the model with little success. A more unconventional approach was taken after the neural net results were consistently lower than the baseline linear regression. Automated Machine Learning, a package which can help find reasonable neural net hyperparameters and architecture. Automated Machine Learning was implemented through a keras package, and helped identify a few differences in model construction which were used. Data was then fed through 50 different trials or unique preconstructed models all with unique parameters and hyperparameters. After the auto-machine learning process was completed, additional adjustments to the model were made. Surprisingly, it was discovered that normalization of the standardized input variables slightly improved results compared to standardized data alone. Regularization for the neural network was implemented by using a combination of dropout layers and early stopping. Three different dropout layers were applied throughout the model, with a 25% rate of dropping inputs. This was done to prevent overfitting.

Likewise, early stopping was implemented. Because the neural networks use the backprop algorithm for fits, the neural networks require multiple iterations known as epochs. Therefore it needs a measure to minimize the loss to the target vector, while preventing overfitting of the model. Early stopping also selects the appropriate number of epochs to be run through the network. For this model, an early stopping was implemented to stop if loss has not decreased within 7 epochs. Additionally, using weight decay with the L2 penalization was originally implemented, but it decreased accuracy on the test data.

The model loss here is determined through Mean Absolute Error (MAE) or also known as the L1 loss function can be seen in equation 4.6. Using MAE here is appropriate, as mean-squared error is highly sensitive to outliers, which as discussed in the *Data* section are extremely prevalent in the H2M ratio target vector. If instead the loss function used mean squared error(MSE), the L2 loss would square the errors which would amplify the outliers effect on loss.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\widehat{y_i} - y_i\right| \tag{4.6}$$

Additionally learning rate needs to be adjusted, the learning rate describes the magnitude in which the stepwise parameter updates. If it is too small it may never reach the minima, and if the learning rate is too large it may miss the true minimum value. The learning rate used in this process was 0.001. After selecting the appropriate parameters, the model loss can be viewed by running through the parameters in figure 4.1. Originally a validation set was used to determine early stopping, but due to data limitation, it was not implemented.

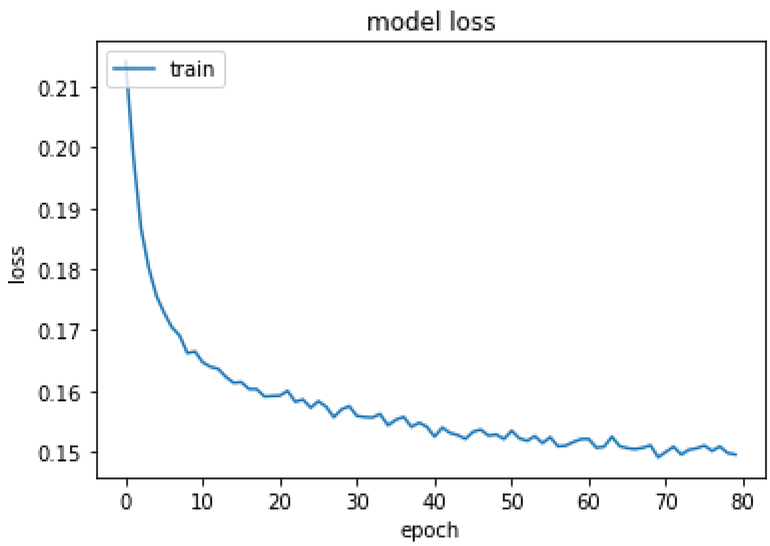

Figure 4.1: Model loss of the neural network over epochs

The final model used is displayed in figure 4.2. This neural network uses 3 hidden layers. The first hidden layer consisted of 32 nodes, 128 nodes in the second hidden layer, and 32 nodes for the third hidden layer. All hidden layers utilized ReLu activation functions. All 32 nodes in the third hidden layer were then fed into the final node representing the output or the regression head, for all 35,084 observations.

```
R squared of model 5 NN is: 0.732
Model: "Model 5 - Regression Head with dropouts"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 input_1 (InputLayer)        [(None, 13)]              0
 multi_category_encoding     (None, 13)                0
 normalization(Normalization) (None, 13)               27
 dense (Dense)               (None, 32)                448
 re_lu (ReLU)                (None, 32)                0
 dropout (Dropout)           (None, 32)                0
 dense_1 (Dense)             (None, 128)               4224
 re_lu_1 (ReLU)              (None, 128)               0
 dropout_1 (Dropout)         (None, 128)               0
 dense_2 (Dense)             (None, 32)                4128
 re_lu_2 (ReLU)              (None, 32)                0
 dropout_2 (Dropout)         (None, 32)                0
 dropout_3 (Dropout)         (None, 32)                0
 regression_head_1 (Dense)   (None, 1)                 33

=================================================================
Total params: 8,860
Trainable params: 8,833
Non-trainable params: 27
_________________________________________________________________
```

Figure 4.2: Final neural network model summary

The results of the neural network often varied. In the loss function in figure 4.1 above, the model was run through 81 epochs. This process was repeated 5 times to account for variation in the results. The model would run through 55-90 epochs. The resulting means of the three loss metrics for all 5 runs will be provided in the results section.

Gradient Boosted Trees

The final algorithm was gradient boosted decision trees. Gradient boosting is an ensemble method which builds predictions by taking the sum of many sequential weak learner models, each built by trying to minimize the loss left over by previous models (Hastie, Friedman, & Tisbshirani, 2009). The weak learners are frequently chosen to be decision trees of low depth, and were selected as the weak learner in this case given their robust characteristics and strong historic performance (Wang, Shi, Lyu, & Deng, 2017). Equation 4.7 shows how the boosted model, $f_m(x)$, is built using the sum of many models, in this case trees T. The small decision trees are built given the variables and a partition, Θ, from equation 4.8. The partition for each weak learner, equation 4.8**,** is defined by minimizing the loss function, squared error in the case of our model, given the unexplained variance from the previous trees. (Hastie, Friedman, & Tibshirani, 2009).

$$f_m(x) = \sum_{m=1}^{M} T(x\,;\,\Theta_m) \tag{4.7}$$

$$\widehat{\Theta_m} = argmin_{\Theta_m} \sum_{i=1}^{N} L(y_i, f_{m-1}(x_i) + T(x_i; \Theta_m)) \quad (4.8)$$

The python package XGBoost was used for implementation of the boosted trees, as it is a widely used library and has strong historical performance (Wang et al. 2017). The hyperparameters considered for tuning were the learning rate, max tree depth, the percentage of variables to sample for each weak learner, whether to apply lasso regularization, and the minimum loss reduction required by the algorithm to warrant splitting a leaf. Learning rate is similar to as explained previously with neural nets, it is the constant value applied to each model to scale down the contributions from each weak learner. The max tree depth is how many splits are allowed for each tree, with a larger depth allowing for more complicated weak learners. Hyperparameters were tuned using 5 fold cross validation from 2304 candidate hyperparameter value combinations with the CVGridSearch package. Different loss functions for the cross validation of hyperparameters were explored, with mean absolute percentage error producing the best result on a holdout segment of data, 10% of training data. Cross validation returned a learning rate of .3, a max tree depth of 6 splits, to not apply lasso regularization, to use all variables for each weak learner, and to set no minimum loss reduction for splitting. The inputs for the XGBoost regressor were the same as the linear regression, with daytype being one-hot encoded and average daily weight for each daytype as the daily clock.

Section 5: Results of Renewable Focused Machine Learning Model

Machine Learning Algorithm Selection

The results of all the models are seen below in table 5.1. The best performing model was the XGBoost model, which outperformed the different neural network models created, along with the lasso, and the linear regression model. Here we will use R-squared to provide the accuracy representation of the regression model. The XGBoost vastly outperformed all other models examined. When fitting the gradient boosted model with hyperparameters selected through cross validation on the test set, provided an R-squared of 0.873. Essentially this means the model accorded for 87.3% of the variation examined in the test set when choosing parameters with cross validation.

Due to the Neural Network's inconsistencies when running the data, the final selected model was run 5 times and averaged for the resulting Test R-squared value of 0.731. Overall the resulting Neural Network was marginally better than the Linear baseline or the lasso model with polynomials and interactions. Although, the results of the Gradient boosted tree provided superior results on all metrics. With the highest R-squared, and lowest MAE and MSE values.

Table 5.1: Renewable energy-focused models resulting metrics

Renewable Focused Models	Training R-squared	Test R-squared	Mean Absolute Error (MAE)	Mean Squared Error (MSE)
Linear	0.6741	0.682	0.1643	0.0594
Lasso	0.7188	0.721	0.1557	0.0521
Neural Network	0.725	0.731	0.1395	0.0504
XGBoost	0.927	0.873	0.1083	0.0236

Results suggest that the XGBoost is the best suited model and will be used as the final model for the rest of the paper. The model then was plotted by feature importance as seen in figure 5.1. The feature metric seen in the plot known as F-score simply sums up how many times each feature is split on. The model included 5069 total splits. The results show oil prices as the most split-on variable, the renewable ratio as the second most split-on variable, with holidays and Monday-Thursday daytype as the least split-on variable. It should be noted that the sum of the feature importance of all day types is 396, and therefore would be the second least split upon variable after COVID.

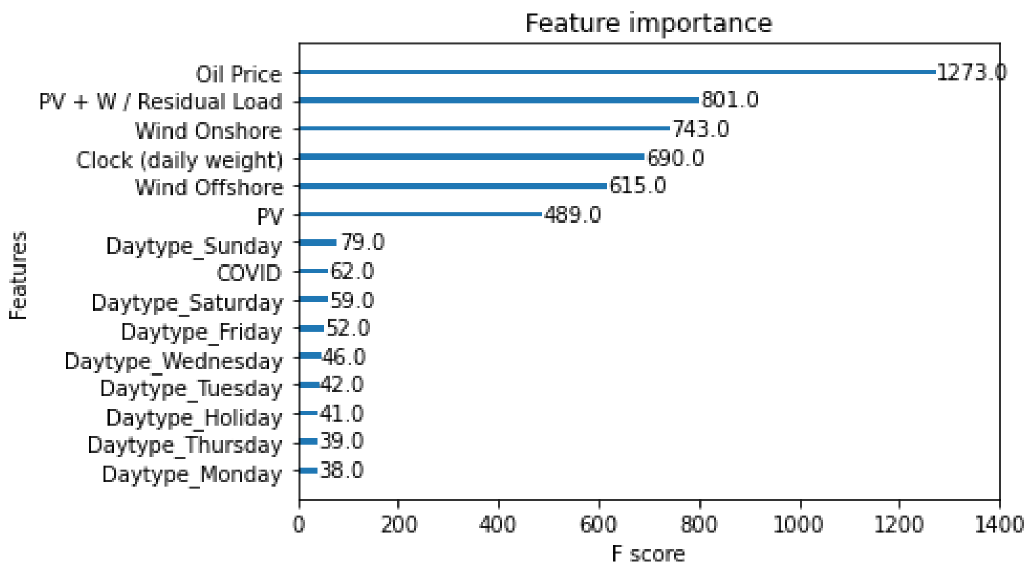

Figure 5.1: Feature importances of input variables for XGBoost

Discussion of Top Performing XGBoost Model

Generally the results of model selection were somewhat surprising given previous research notably by Green (2014). Although the superior results of utilizing XGBoost align with other similar papers, such as Wang et al. (2017) where electricity consumption prediction was best estimated using XGBoost in combination with discrete wavelet transform, when comparing alternative models. It should be noted that due to the time limitation within the dataset, may hold in part the reasoning behind the neural network's shortcoming.

Overall the gradient boosted tree model provided the strongest result. As seen in Ballester, and Furió (2014) and Beolet, de Jong, & Enev (2014), renewable energy sources were observed to be important features. Wind onshore, wind offshore, photovoltaic energy and the renewable ratio all appeared to be significant variables according to the feature importance figure 5.1.

Equation 5.1 calculates the splits calculated from renewable energy based variables and will be referred to as the Percentage of Renewable Splits (PRS_s) , with s symbolizing the number of splits. The variable renewable generation variables themselves (wind onshore, wind offshore, and photovoltaics) represented a total of 1,847 splits. This equation also includes the renewable ratio(RR) where renewable split decisions sum to 2648. 2648 out of 5069 total splits are based on renewable energy based variables using this model. This makes up 52.2% of all split decisions.

$$PRS_s = \frac{\sum_{i=1}^{3}\left(Won_s + Woff_s + PV_s + RR_s\right)}{\sum_{i=1}^{N}\left(X_s\right)} * 100 \qquad (5.1)$$

It is understood that the total number of splits is not a true statistic for percentage influence that the variable has on the model's outcome. Given that limitation, it is one of the best measures to use for calculating variable importance in a gradient boosted tree model. This statistic produced in equation 5.1 also assumes that each split decision accounts for an equal significance for the estimated daily shape, which again is a known bias. The percentage of renewable splits resulted in a value of 52.2%, which can be applied as an estimate to claim 52.2% of the model arises from renewable energy to explain the hourly profiles in the day-ahead market. This is a significant result which gives merit for renewable's focus in this paper.

One note of interest is that the F score for the day of the week categorical variable (representing the intra-weekly clock) accounted for 396 splits, which is low. Overall the time of day (intra-daily clock variable) accounted for 690 splits and is nearly twice as prevalent. Overall Sunday and saturday proved to be the most split on days, as they have the most unique average average shapes as seen in figure 3.8 in the *Data* section.

One of the most noteworthy results is from the significance of the feature engineered renewable ratio variable represented as (PV + W/ Residual load) in figure 5.1. This variable has the second most splits in the renewable focused gradient boosted model, with 801 splits. It is surprisingly split upon more often than the intra-daily clock variable, and all other renewable energy generation sources, including wind onshore.

Wind offshore surprisingly had a larger influence on the number of splits than photovoltaic generation. This is surprising due to the larger installed net power generation capacity that photovoltaic energy accounts for in Germany (Appunn, Haas & Wettengel, 2021), as well lower actual generation in the data. Perhaps offshore wind is a more variable energy source than solar energy and requires more splits to account for the all day generation, along with increased volatility fluctuation.

The most significant variable according to the feature selection plot are oil prices. When comparing oil prices to the H2M ratio along with the COVID lockdowns, you can view the correlation in figure 5.2. The connection with standardized oil prices and the standardized H2M ratio when the oil price collapsed between March 2020 to May 2020 depicts a notable relationship between oil prices and the H2M ratio. This is clear as the H2M ratio and oil prices experienced their lowest values. It can be reasonable to assume oil prices incorporate the account for COVID-19 lockdowns as lockdowns were worldwide during this time. Although in figure 5.2, the second national lockdown in Germany did not appear to have a significant effect on the H2M ratio. When closely examining oil prices above the standardized price of 1, the standardized H2M ratio registers nearly no negative standardized prices. This proposes a strong connection between higher oil prices (at least of values close to 1.0) and the stability of H2M-ratio, and in turn day-ahead spot prices. This plot helps to signify why oil prices had the largest feature importance F-score, and likely helped the renewable focused gradient boosted tree model account for the presence of outliers within the model.

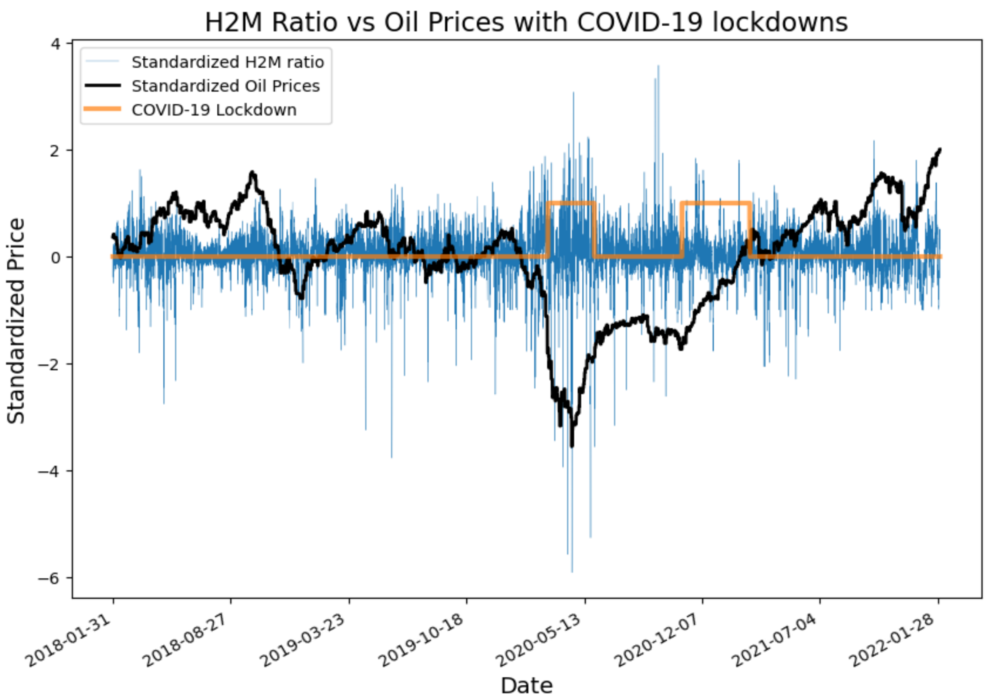

Figure 5.2: Spot price H2M (Standardized), Oil Price (Standardized), and COVID Lockdown

The oil price dip in 2020 into negative prices depicted the crisis related to the COVID pandemic, as displayed in figure 5.2. Additionally, the recent increase in oil prices from mid-2021 represents the end of lockdowns, and the return to normality in the world's economy.

The COVID-19 variable did not appear to be overly relevant for modeling, while oil prices had a more significant role when predicting spot price. It could be the case that oil prices already partially represent the latent covid variable, while also containing other significant macroeconomic variables as discussed by Jones & Leiby (1996).

Analysis of Three Characteristic Periods

The following plots will incorporate 2 different model predictions, along with the real values of the day-ahead spot prices expressed by the H2M ratio target vector. The first model will represent the final renewable focused XGBoost model that was described in the *Methodology* section and was the best model created for this research. The second model is a linear regression model which includes all generation and consumption variables found on SMARD.de.

The second model will be a completely new model, which only serves the purpose as a reference for the renewable focused gradient boosted model produced in this paper. This model will serve as an example of a standard model, which incorporates all energy generation and consumption statistics from SMARD.de. This includes conventional sources of electricity and shall be referred to as the 'conventional linear model'. This model will also be given advantages to improve its performance, and therefore present as a reasonable model that could be used in the industry today. Firstly, this linear model was also given the advantage of being overfit, by not being subjected to a training and test set, and trained on all data which provides it with more accurate predicted values on this set. Secondly, this model includes oil prices, and has been given polynomial transformations for all variables and includes variable interactions, similar to the lasso model as described in the *Methodology* section.

This conventional model contains 125 variables, compared to just 15 variables in the renewable focused gradient boosted model. This conventional model primarily focuses on conventional generation, while still containing renewable generation. This linear model does not include the renewable ratio, or the covid variable as inputs. It can therefore be viewed as a standard model, without the focus on renewable generation, and without incorporating the COVID-19 lockdown shock periods. This linear model was made with unrealistic assumptions in terms of ex ante information and performed well on the data, resulting with a training R-squared value of 0.781, an MAE of 0.139, and an MSE of 0.042. This model's metrics are better than the linear, lasso and neural network's resulting metrics.

There will be three plots generated to show how the renewable focused gradient boosted model performs in different scenarios. First plot will be a standard period where the renewable ratio is below 1, the time before COVID lockdowns in 2018, and when oil prices were steady. The second plot will represent a period of volatility during the first COVID lockdown, when oil prices and spot prices are below zero. The last plot represents a day when the renewable ratio is exceptionally high and should be a good representation of how the model will adjust to future generation of renewable energy.

The first example is displayed in figure 5.3, and was chosen as it was in a period where the renewable ratio was below 1, before the COVID lockdowns, and while oil prices were at a standard level. In this plot below you can see that the daily shape curve represented from both models provides decent predicted shapes, but still both have inaccuracies. It should be noted that the renewable focused gradient boosted model and conventional model both overestimate and underestimate prices with a similar margin of error, with the gradient boosted model predicting higher than average prices. This could be the effect as prices in late 2021and 2022 were much higher while under similar conditions.

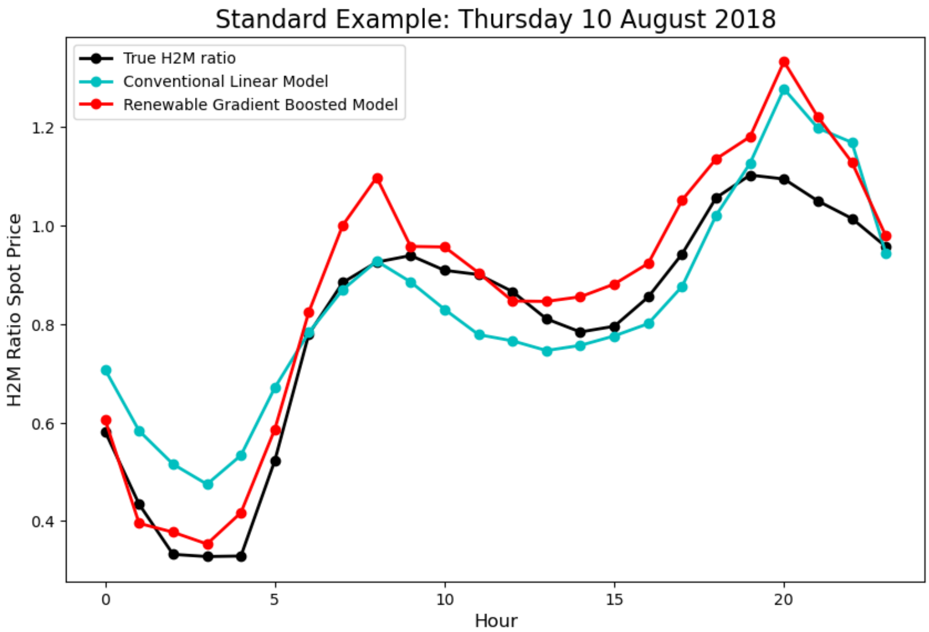

Figure 5.3: Model performances vs observed H2M prices on a normal day

The second example period is shown below in figure 5.4. This day was chosen as it is in the middle of the first COVID-19 lockdown. The resulting period was also chosen as it had close to the lowest day-ahead spot prices in Germany, where the H2M ratio was below -4.0 for 6 hours, and was during the period when oil price were very low at 9.87 US dollars a barrel within the EU compared to the mean during the 4 year period being 62.37. Additionally the renewable ratio was above 1.0 all day, and spiked mid-day from high Photovoltaic energy production. From 10:00 to 19:00 the renewable focused gradient boosted model had near perfect predictions, while the conventional linear model provided very poor estimated prices, missing values by a value of around 3.0 for 6 hours straight, which is a substantial error. Additionally, it should be noted that even with the huge midday dip, the renewable focused XGBoost model also provided more accurate predictions throughout the entire day. It can be suggested that the incorporation of the covid variable, alongside the renewable ratio provided a much more accurate prediction by the XGBoost model, which provided extremely accurate results.

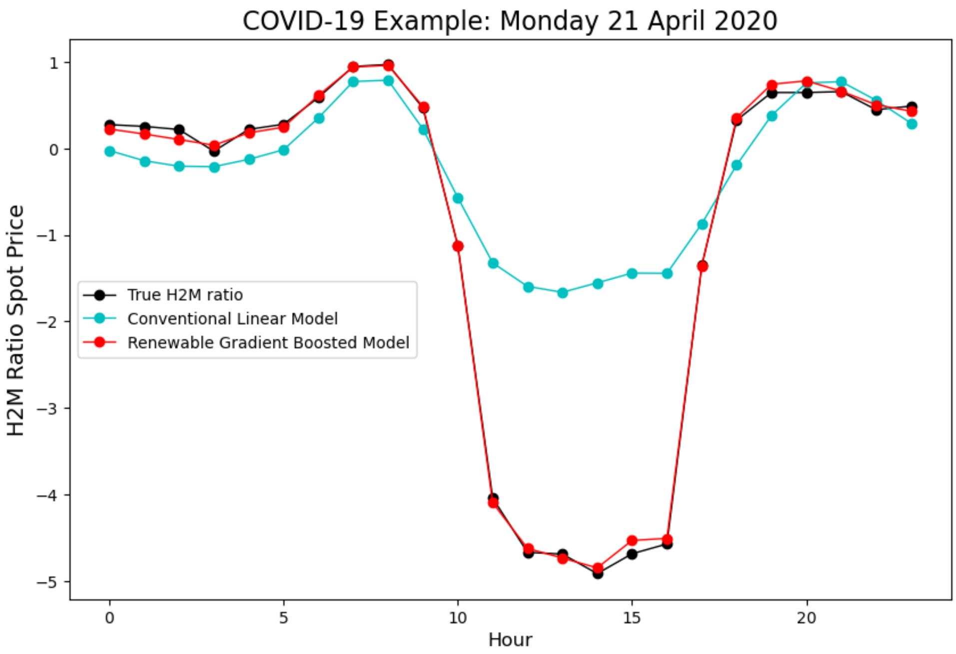

Figure 5.4: Model performances vs observed H2M prices during COVID-19 lockdown

Figure 5.5 displays the high renewable ratio effect on the models' day-ahead shape curve. January 30 2022 was selected because it contained the most hours with the renewable ratio over a value of 3.0. These are hours where renewable energy generation was three times greater than conventional generation, this was primarily due to the high amount of onshore wind generation in the morning of the day. The differing results of the renewable focused gradient boosted model again creates a better prediction of spot prices comparative to the conventional model.

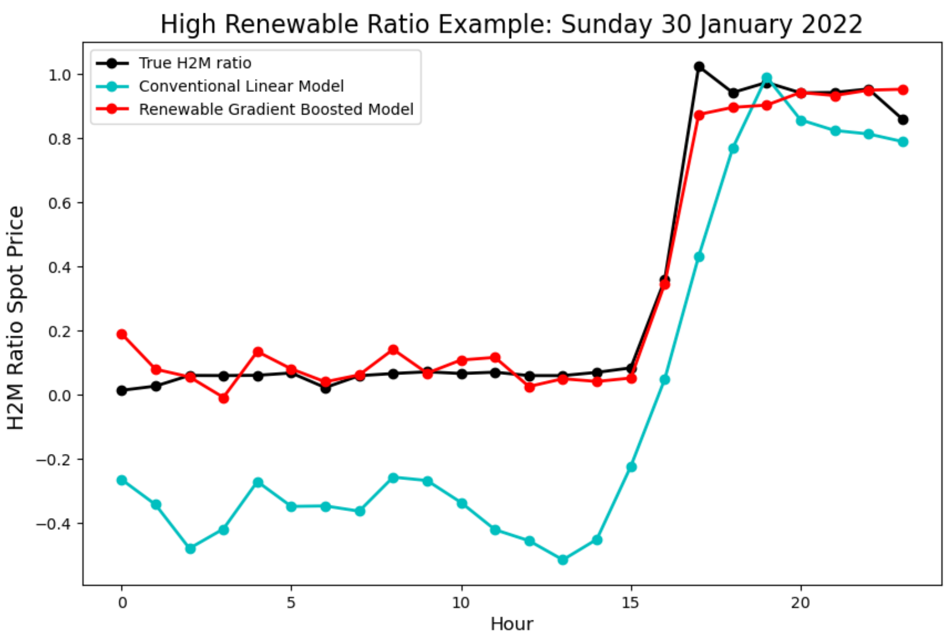

Figure 5.5: Model performances vs observed H2M prices during high renewable production

Figure 5.6 includes the renewable ratio combined with the high renewable ratio example. From 00:00 until 15:00 the renewable ratio registered values between 3.6-7.8, which are very large outliers in the dataset. The ratio rapidly decreased to less than 1.0 after 18:00. You can see in the results that the renewable focused gradient boosted tree models were very accurate during the hours containing a high renewable ratio, while the conventional linear model underestimated the price. Only as the renewable ratio dipped below 1.0 at 18:00 is when the conventional model converged back to accurately predicting the H2M ratio. This plot demonstrates that the renewable focused gradient boosted model does well to adapt to an extreme influx of VRE generation. The model accurately estimates the shape of the day-ahead spot prices curve during these unusual conditions.

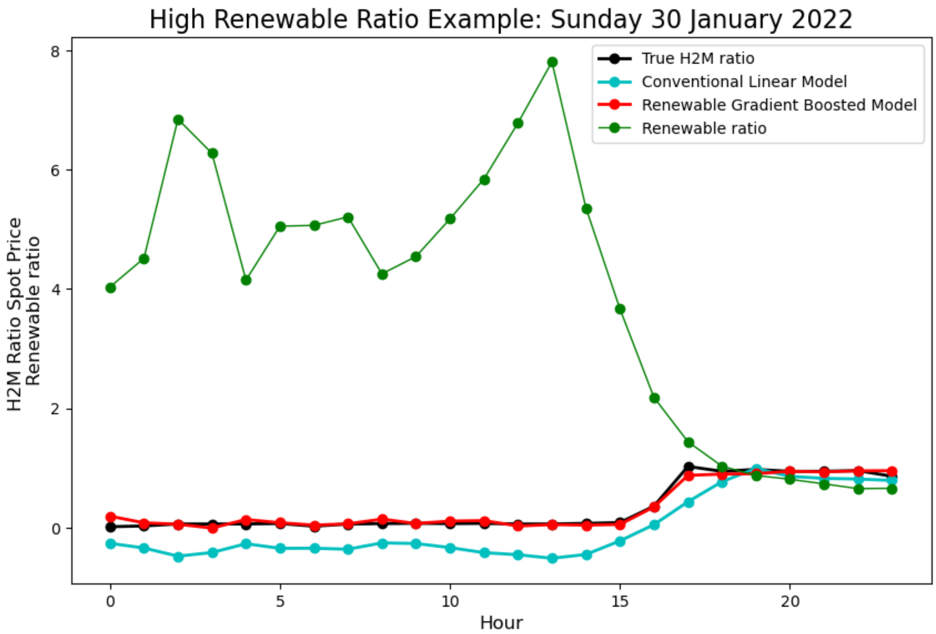

Figure 5.6: Model performances vs observed H2M prices during high renewable production with renewable production included

The resulting estimates by the renewable focused extreme gradient boosted tree model showed substantial differences than the conventional model. The focus on selecting renewables appeared to not take away too much information during average days, as seen in the first example. There is also evidence that it has superior performance in estimation during periods of COVID lockdown, and when VRE generation is substantial. Additionally, it should be noted that the importance of VRE in the feature importance figures suggests Day-Ahead energy traders in the German market should pay special attention to the variables and their interactions. A shifting market should also suggest a change toward limited conventional generation inclusion, focusing more on renewables and the proportion of renewable energy generation, inclusion of exterior variables to the market such as European oil prices, and a lockdown variable to account for external influences to the market, while still including seasonality, intra-weekly and intra-daily patterns.

Section 6: Limitations and Future Research

Limitations

A significant limitation in this research is that actual historical measurements of electricity generation from VRE were used inplace of forecasted generation. This was done to insulate the hourly profiles from the possible bias included in the unknown models used to predict VRE electricity generation. Implementation of the hourly profiles from this research to predict future spot prices after being calibrated would require forecasted data on renewable production, which could result in different results and effectiveness than are reported here. While models constructed for this research were agnostic to VRE forecasting techniques, practical implementation would not have this luxury.

An additional limitation is how renewable energy is impacted by unforeseen events in erratic ways. Market shocks can be studied ex-post, but each shock can be expected to impact the market in different ways and no one “shock variable” can be expected to account for all current events. For example, a “shock” scenario like COVID was implemented into our data which showed a decrease in price, but that variable could not be used to account for the Russian invasion of Ukraine. The Russian invasion has increased energy prices significantly, but could not be accounted for at the time of this research. It could be expected that, concerning the invasion of Ukraine, political leaders will make further decisions with substantial impact on energy markets that would be all but unknowable ex ante. Similar shocks will always impact the reliability of prediction and serve as a limitation for the methodology included in this research.

Another main limitation of this research, and for other similar papers involving the electricity market, is the time window definition. As stated in the literary review, technological learning rates alongside the increasing impact that renewables have on the market creates a tradeoff with long windows of data being less representative of future conditions, but shorter windows being too limited for models to gain sufficient training. This can be explored by evaluating the number of different temporal combinations, the day of week, month, and hour combinations. Each combination is expected to act with distinct characteristics given market conditions. 35084 observations over 7 days a week over 24 hours a day leaves 209 unique sets of weekly data and having only 209 sets could present a data limitation issue.

Future Research

It should be noted that the next step for application of the model would be configuration of the shape vector of the renewable focused gradient boosted tree model with respect to spot prices. As per Green (2014) and Burger, Graeber, and Schindlmayr (2008), the same process could be made to use a linear scaling method, and after this would be completed could it be applied. Although the purpose of the thesis was generating the hourly price forward curve vector, therefore this excludes the calibration and application of the model. So application of the model should be the next step in future research.

Research on the causal link between oil price and electricity price in the current market would contribute to the understanding of why oil price was identified as so significant in the gradient boosted tree model. As stated in the limitations, oil prices have increased to historically high levels. As stated in the Kost et al.(2021), the renewable LCOE has rapidly decreased. Research could be conducted to see how oil prices affect electricity prices long term as high oil prices push people to renewable sources of energy. The resulting impact of this push could be positively correlated to price as it increases demand for electricity, or the impact of the push towards renewables could be negatively correlated to energy price, as the decreasing cost of renewables by means of increased economies of scale, alongside increasing the technological learning rates of renewables can cause energy to be cheaper in the long-run. Research on the relationship between oil price and renewable production could be important to understanding the day-ahead energy markets in the future.

A potentially substantial finding from the research conducted for this paper was how the second national lockdown in Germany had no significant effect on the hour to monthly average ratio of spot price. Future research should be conducted on whether this result is confounded by the incorporation of COVID lockdown behaviors into other model variables. A variable outside of the COVID categorical variable accounting for lockdown behaviors would be unlikely as year was not provided in the modeling process, making it impossible for models to differentiate a month during a COVID lockdown from that month without COVID lockdown in the absence of the COVID categorical. However, it could also be that the COVID dates collected were poor indications of observed lockdown behavior, with local lockdowns and non-observance of restrictions confounding the results in this paper. More comprehensive research of this topic is needed to make statements about this relationship.

A compelling suggestion for future research would be to construct a Convolutional Neural Network (CNN). This could be accomplished by using the hour/sum to unity variable for the dependent variable which was previously described in the data section. To implement this, each day would be changed into panels. The resulting panels would have 24 hours on the y-axis, with the independent variables on the x-axis. This represents a similarity to an image, which convolutional neural networks are the most commonly implemented for. By using CNNs, it would output 1461 daily shapes as the target vectors which represent the days used in the dataset of this paper. This was a recommendation by Rikard Green for the paper, although it was not implemented due to time constraints. If subsequent research is to be conducted, utilizing a CNN for modeling is among the first recommendations.

Another avenue for an improvement of modeling spot price and volatility would be to incorporate Urgent Market Messages (UMM), which are planned or unplanned adjustments to production like a temporary shutdown of a gas factory for maintenance. Although the generation data already contains much of the UMM information, explicit inclusion of UMMs could help account for some variability missed. Incorporating localized shutdowns of different energy sources is an example of where UMMs and an input could produce gains in predictive accuracy. More localized effects from the day-ahead curves could be done for particular regions of Germany, by modeling their specific demand, and perhaps could be needed by regional requirements.

Another important feature of exploration could be to incorporate lagged time series variables into the model. The overall purpose was for hourly price forward curves to not be shaped by a time series regression model as described in the *Literature Review*. Implementing lagged time series variables into the model could be a focus for future research. An hourly lag of the renewable ratio could affect the hourly profiles, while a lagged oil price variable by a length of a week or month could also be influential.

Technological learning rates have been discussed throughout this paper. Changing renewable energy proportions within Germany's electricity supply could require the algorithm to be switched to account for new patterns or models to be updated with more recent data. For example, this would be true if large technological innovations are made within renewable technology or other electrical sources. Substantial innovations being made within electricity storage technology could result in electricity switching its behavior away from a flow commodity and would require extensive research to adjust from the contemporary methodology (Peremans, 2018). If the economic viability of electricity storage increases along with technology improvements for storage of electricity, the shape curves could be entirely different from how they appear currently. Therefore microgrid configurations as performed in Peremans (2018) could provide an avenue for relevant research in the future.

The models developed for this research were constructed with the continued increase in the share of renewable energies in mind. The proportions of Germany's electricity market will continue to shift, and therefore intra-daily, and intra-weekly shapes will adjust. Shifting of the energy market will require adjustments in the parameters and the introduction of other variables exogenous to the energy market, if modeling with future data. Another factor over time, geopolitical events, like the Russian invasion of Ukraine, will have unpredictable consequences on energy market behavior and require continuous research. Exemplified by the abandonment of the 11 billion euro investment in the Nord Stream 2 pipeline, Russian military action has drastically changed Germany's energy and electricity market, most likely having an effect on the shape curve when the renewable ratio value is very low (Silverstein, 2022). New research that accounts for developing geopolitical decisions would likely be critical to understanding the macro shifts unaccounted for in the research for this paper.

If these hourly profiles were to be used for the calibration process, for the specific purpose to predict day-ahead prices in the upcoming years, it would be suggested that the model would incorporate future investments of renewable energy. Data on such investments, along with forecasted investments are available within the Kost et al. (2021) which specifies market development and forecast renewable energy investment. This also categorizes investment separately for both wind power and photovoltaic energy generation (Kost et al. 2021). As the model has better been adjusted for future generation, this input inclusion is recommended to get a better model especially for data years into the future.

Section 7: Conclusion

Throughout this paper we have examined the effect of four different machine learning algorithms to be applied to create the shapes of the hourly price forward curves for the German market. All four algorithms tested only contain VRE generation and renewable proportion inputs, clock inputs to represent demand, and exogenous inputs. The paper aimed to make the best machine learning regression model to create a renewable energy-based shape of the hourly price forward curve.

The paper's core attempted to answer the question of whether VRE-related generation alone, clock inputs, and further exogenous inputs can significantly capture volatility in the hourly shape profiles of the German EPEX Spot market to model hourly price forward curves. The resulting final gradient boosted model only incorporated 15 variables and outperformed the conventional linear model, which included 125 variables. Our renewable-based model outperformed the overfit conventional model, resulting in a higher R-squared, and lower MAE and MSE values. The model also proved to be significantly more reliable in the volatile situation examples, as seen in two examples in the *Analysis of Three Characteristic Periods* subsection. The model also shows significance in the renewable ratio variable. The ratio had a significant number of splits in the gradient boosting model. It showed a significant difference in the last period examined in the *Analysis of Three Characteristic Periods* subsection.

The output metrics and hourly price forward curves of the resulting gradient boosted trees model captured the H2M ratio well and, by extension, could be used to predict day-ahead spot prices well. As stated, this model currently needs adaptation to current situations as the Russian invasion of Ukraine would require another exogenous input into the gradient boosted tree model. The hourly day-ahead shape could then be calibrated to the German EPEX SPOT day-ahead market spot prices. This model could be implemented to predict future hourly profiles when renewable production forecasts are provided. The calibrated model could then be used as an hourly price forward curve for day-ahead traders in the German EPEX SPOT day-ahead market.

References:

Abu-Rayash, A., & Dincer, I. (2020). Analysis of the Electricity Demand Trends Amidst the Covid-19 Coronavirus Pandemic. *Energy Research & Social Science*, vol. 68, Available online: https://doi.org/10.1016/j.erss.2020.101682 [Accessed 12 April 2022]

Appunn K., Haas Y., & Wettengel J. (2021). Germany's Energy Consumption and Power Mix in Charts, *Clean Energy Wire,* Available online: https://www.cleanenergywire.org/factsheets/germanys-energy-consumption-and-power-mix-charts [Accessed 8 April 2022]

Ballester, C., & Furió, D. (2015). Effects of Renewables on the Stylized Facts of Electricity Prices, *Renewable and Sustainable Energy Reviews*, vol. 52, pp.1596-1609, Available online: https://doi.org/10.1016/j.rser.2015.07.168 [Accessed 12 May 2022]

BBC. (2020). Coronavirus: German Alpine region goes into lockdown, British Broadcasting Corporation, Available online: https://www.bbc.com/news/world-europe-54618827 [Accessed 14 April 2022]

Beolet, V., de Jong, C. & Enev, E. (2014). Improved Hourly Shaping Using Renewable Production Information, *KYOS Energy Consulting BV*

Bofinger, S., Saint-Drenan, Y., Amany von Oehsen, A. Norman Gerhardt, N., Sterner, M., & Rohrig, K. (2009). Dynamische Simulation der Stromversorgung in Deutschland nach dem Ausbauszenario der Erneuerbaren-Energien-Branche, *Fraunhofer Institute for Solar Energy Systems ISE*, Available online: https://www.bee-ev.de/fileadmin/Publikationen/Studien/100119_BEE_IWES-Simulation_Stromversorgung2020_Endbericht.pdf [Accessed 10 April 2022]

Bord, B. (2022). Workalender. Available online: https://pypi.org/project/workalendar/ [Accessed 25 April 2022]

Burger, M., Graeber, B. and Schindlmayr, G. (2008), *Managing Energy Risk: An Integrated View on Power and Other Energy Markets*, Wiley Finance, West Sussex

Burger, M., Graeber, B. & Schindlmayr, G. (2014). *Managing Energy Risk: A practical guide for risk management in power, gas and other energy markets*, New York, Wiley

Crispin, D., & Jacobsson, C. (2007). General Forward Curve Specification: An optimisation approach with price shape statistics, Working paper, E.ON Sales and Trading

Deisenroth, M. P., Faisal, A. A., & Ong, C. S. (2020). Mathematics for Machine Learning, pp.370-392, Cambridge: Cambridge University Press

DeLuca, A. (2018). The Race to Renewables: A comparative study of the energy market in Germany and the United States, *Oregon Review of International Law*, vol. 19, no. 1, pp.259–288, Available online:

https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/porril19&section=9 [Accessed 8 May 2022]

de Jong, C., van Dijken, H., & Enev, E. (2013). How Renewables Shape the Future, *KYOS Energy Consulting*

DW News. (2020). Germany Won't Lift Coronavirus Restrictions Before April 20: Chancellory, *DW News,* Available online: https://www.dw.com/en/germany-wont-lift-coronavirus-restrictions-before-april-20-chancellory/a-52947201 [Accessed 14 April 2022]

DW News. (2021). Coronavirus: Germany extends lockdown with plan to relax restrictions. *DW News,* Available online: https://www.dw.com/en/coronavirus-germany-extends-lockdown-with-plan-to-relax-restrictions/a-56763824 [Accessed 14 April 2022]

Duffy, A., Hand, M., Wiser, R., Lantz, E., Riva, A., Berkhout, V., Stenkvist, M., Weir, D., & Lacal-Arántegui, R. (2020). Land-based Wind Energy Cost Trends in Germany, Denmark, Ireland, Norway, Sweden and the United States, *Applied Energy*, vol. 277. Available online: https://doi.org/10.1016/j.apenergy.2020.114777 [Accessed 16 April 2022]

Ettlin, A. (2019). ‘Improved Hourly Price Forward Curve Forecasting Based on Future Prices for Hydropower Production’, 2019 16th International Conference on the European Energy Market (EEM), European Energy Market (EEM), 2019 16th International Conference, pp.1–5. doi: 10.1109/EEM.2019.8916338, Available online: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8916338 [Accessed 05 May 2022]

Fonti, V., & Belitser, E. (2017). Feature Selection Using Lasso, *VU Amsterdam Research Paper in Business Analytics*, vol 30, pp.1-25.

Frömmel, M., Han, X., & Kratochvil, S. (2014). Modeling the Daily Electricity Price Volatility with Realized Measures. *Energy Economics*, vol. 44, pp.492-502, Available online: https://doi.org/10.1016/j.eneco.2014.03.001 [Accessed: April 12 2022]

Genoese, F., Genoese, M., & Wietschel, M. (2010). Occurrence of Negative Prices on the German Spot Market for Electricity and their Influence on Balancing Power Markets. 2010 7th International Conference on the European Energy Market, Energy Market (EEM), Available online: https://doi-org.ludwig.lub.lu.se/10.1109/EEM.2010.5558703 [Accessed: 18 May 2022]

González-Aparicio, I. & Zucker, A. (2015). Impact Of Wind Power Uncertainty Forecasting on the Market Integration Of Wind Energy In Spain. *Applied Energy*, vol. 159, pp.334-349, Available online: https://doi.org/10.1016/j.apenergy.2015.08.104 [Accessed: 14 May 2022]

Green, R. (2014). A Power Market Forward Curve with Hydrology Dependence - An Approach based on Artificial Neural Networks. Available online:

https://search-ebscohost-com.ludwig.lub.lu.se/login.aspx?direct=true&AuthType=ip,uid&db=edsswe&AN=edsswe.oai.lup.lub.lu.se.d2094c10.f8c3.467f.b7d4.14528c6a9b4d&site=eds-live&scope=site [Accessed: 16 February 2022].

Hansen, K., Mathiesen, B.V. and Skov, I.R. (2019). Full Energy System Transition Towards 100% Renewable Energy in Germany in 2050. *Renewable and Sustainable Energy Reviews*, vol 102, pp.1-13.

Hastie, T., Tibshirani, R. & Friedman, J. (2009). *The Elements of Statistical Learning: Data mining, inference, and prediction*, 2nd edn., New York: Springer.

Hickey, E., Loomis, D., & Mohammadi, H. (2012). Forecasting Hourly Electricity Prices Using ARMAX–GARCH models: An application to MISO hubs. *Energy Economics*. vol. 34, Issue 1, pp.307-315, Available online: https://doi.org/10.1016/j.eneco.2011.11.011 [Accessed 26 April 2022]

Holm, D., & McIntosh, J. (2008). Renewable Energy: The future for the developing world. *Renewable Energy Focus*. vol. 9, no. 1, pp.56-61, Available online: https://doi.org/10.1016/S1471-0846(08)70027-1 [Accessed 22 April 2022]

Jones, D.W. & Leiby, P.N. (1996). The Macroeconomic Impacts of Oil Price Shocks: A review of literature and issues. *Oak Ridge National Laboratory*. vol. 25(2), pp.1-32.

Kost, C. (2021). Levelized Cost of Electricity: Renewables clearly superior to conventional power plants due to rising CO2 prices. *Fraunhofer Institute for Solar Energy Systems ISE*. Available online: https://www.ise.fraunhofer.de/en/press-media/press-releases/2021/levelized-cost-of-electricity-renewables-clearly-superior-to-conventional-power-plants-due-to-rising-co2-prices.html [Accessed 18 April 2022]

Kost, C., Shammugam, S., Fluri, V., Peper, D.,Jülch V., Nyugen H.T., & Schlegl, T. (2018). Levelized Cost Of Electricity Renewable Energy Technologies. *Fraunhofer Institute for Solar Energy Systems ISE*. Available online: https://www.ise.fraunhofer.de/content/dam/ise/en/documents/publications/studies/EN2018_Fraunhofer-ISE_LCOE_Renewable_Energy_Technologies.pdf [Accessed 15 May 2022]

Kost, C., Shammugam, S., Fluri, V., Peper, D., Memar A.D., & Schlegl, T. (2021). Levelized Cost Of Electricity Renewable Energy Technologies. *Fraunhofer Institute for Solar Energy Systems ISE*. Available online: https://www.ise.fraunhofer.de/content/dam/ise/en/documents/publications/studies/EN2021_Fraunhofer-ISE_LCOE_Renewable_Energy_Technologies.pdf [Accessed 10 April 2022]

Ketterer, J. (2014). The Impact Of Wind Power Generation On The Electricity Price In Germany. *Energy Economics*. vol 44., pp.270-280. Available online: https://doi.org/10.1016/j.eneco.2014.04.003 [Accessed 5 May 2022]

Matschoss, P., Bayer, B., Thomas, H., & Marian, A. (2019). The German Incentive Regulation And Its Practical Impact On The Grid Integration Of Renewable Energy Systems, *Renewable Energy*, Elsevier, vol. 134(C), pp.727-738.

Miller, S. (2020). The Secret to Germany's COVID-19 Success: Angela Merkel Is a Scientist. *The Atlantic.* Available online: https://www.theatlantic.com/international/archive/2020/04/angela-merkel-germany-coronavirus-pandemic/610225/ [Accessed 14 April 2022]

Milis, K., Peremans, H. and Van Passel, S., (2018). Steering the Adoption of Battery Storage through Electricity Tariff Design, *Renewable and Sustainable Energy Reviews*, vol. *98*, pp.125-139

Rintamäki, T., Siddiqui, A., & Salo, A. (2017). Does Renewable Energy Generation Decrease The Volatility Of Electricity Prices? An analysis of Denmark and Germany. *Energy Economics*, vol. 62, pp.270-282, Available online: https://doi.org/10.1016/j.eneco.2016.12.019 [Accessed 6 May 2022]

Schill, W. (2014). Residual Load, Renewable Surplus Generation and Storage Requirements in Germany. *Energy Policy*, vol. 73, pp 65-79, Available online: https://doi.org/10.1016/j.enpol.2014.05.032 [Accessed 12 May 2022]

Seythal, T., & Carrel P. (2020). German Social Distancing will be Extended to May 10, Reuters*,* Available online: https://www.reuters.com/article/us-health-coronavirus-germany-braun-idUSKBN22C1C1 [Accessed 14 April 2022]

Silverstein, K. (2022). Russia's Invasion Kills Nord Stream 2 and Gives Rise to Renewables. Forbes. Available online: https://www.forbes.com/sites/kensilverstein/2022/05/15/russias-invasion-kills-nord-stream-2-and-gives-rise-to-renewables/?sh=24a60ddd73b0 [Accessed May 16 2022].

Sky News. (2020). Coronavirus: France and Germany have announced second national lockdowns, Sky News, Available online: https://news.sky.com/story/coronavirus-germany-to-enter-four-week-lockdown-from-november-chancellor-merkel-confirms-12117056 [Accessed 14 April 2022]

SMARD.de. (2022). Strommarktdaten, German Electricity Market Data, SMARD, Available online: https://www.smard.de/en/downloadcenter/download-market-data#!?downloadAttributes=%7B%22selectedCategory%22:false,%22selectedSubCategory%22:false,%22selectedRegion%22:false,%22from%22:1652479200000,%22to%22:1653429599999,%22selectedFileType%22:false%7D [Accessed 16 February 2022]

Umweltbundesamt. (2022). Renewable Energies in Figures. Available online: https://www.umweltbundesamt.de/en/topics/climate-energy/renewable-energies/renewable-energies-in-figures [Accessed 10 April 2022]

US Federal Reserve. (2022). Download Data for Crude Oil Prices: Brent - Europe. Available online: https://alfred.stlouisfed.org/series/downloaddata?seid=DCOILBRENTEU [Accessed 7 April 2022]

Wang, W., Shi, Y., Lyu, G. and Deng, W. (2017). Electricity Consumption Prediction Using Xgboost Based On Discrete Wavelet Transform. 2017 2nd International Conference on Artificial Intelligence and Engineering Applications (AIEA 2017), Available online: https://scholar.archive.org/work/ry7aqfglhzb3nfhaikarc3vf3q/access/wayback/http://dpi-proceedings.com/index.php/dtcse/article/download/15003/14516. [Accessed 20 May 2022]

Xu, T., Gao, W., Qian, F., & Li, Y. (2022). The Implementation Limitation of Variable Renewable Energies and its Impacts On The Public Power Grid, Energy. vol. 239, Part A. Available online: https://doi.org/10.1016/j.energy.2021.121992 [Accessed 12 April 2022]