Modelling Seasonalities of HPFCs Using a Parametric Approach

Department of Mathematical Statistics
Lund University

LUND
UNIVERSITY

Lucas Svantesson and Reza Rastegar

2019-06-10

Abstract

Electricity differs from other commodities in that it cannot be stored. This non-storability characteristic results in traditional pricing methods for commodities not being applicable for electricity. An alternative pricing method is therefore needed and the solution is the Hourly Price Forward Curve (HPFC). The HPFC essentially gives the prices as of today for the delivery of electricity at each hour in the future. It is generally constructed in two steps. The first step involves estimating the shape vector, which is a vector of hourly weights reflecting all seasonalities and forward-looking information in the electricity spot price. The second step calibrates the shape vector to futures products in order to make it arbitrage free.

In this thesis, we have exclusively studied the first step of the HPFC construction. Specifically, we have modelled the month-to-hour ratios of the shape vector, i.e. a vector of hourly weights normalized for every month. This paper aims to explore the possibilities of modelling the month-to-hour ratios using parametric models with external inputs. This is done by implementing regression models using polynomial and Fourier basis functions, which is thereafter further developed with the addition of a Kalman filter and regularization techniques. The study covers the Nord Pool electricity market and it is conducted using data from E.ON and Nord Pool. It is concluded that parametric models with external inputs are well-suited for constructing a shape vector. It is successful in modelling the intra-yearly, intra-weekly and intra-daily seasonalities and shows robustness against extremities. However, difficulties arose in adequately modelling the summer months and the morning levels.

Keywords: Power Markets, Hourly Price Forward Curves, Seasonality, Electricity Spot Price

Acknowledgements

This thesis was written in cooperation with E.ON Energy Economics, Malmö, Sweden. We would like to express our deep gratitude towards our supervisors at E.ON; Rikard Green, Olivia Nabbosa and Anton Levin. Thank you for your continuous support and immense knowledge sharing and for giving us the opportunity to write this thesis. We would also like to thank the German team for their valuable insights and shared resources; Philippe Witman, Michael Repnikov and Annika Zipperer.

Furthermore, we would like to thank our university supervisor Prof. Erik Lindström for his great guidance and mentoring of our thesis.

Finally, a warm thank you to all our friends and family, who have supported us throughout our studies.

Lund, June 2019

Lucas Svantesson & Reza Rastegar

Contents

Chapter 1

Introduction

The gradual liberalization of the European electricity markets started in the early 1990s. Until then, the European electricity sector was organized as a regulated monopoly. In each country, one or more vertically integrated companies were solely responsible for generation, transmission, distribution and supply of electricity. The liberalization has been an incremental process, with policies spread over three legislative packages (European Parliament, 2016). This recent deregulation offers benefits to both providers and consumers, which stems from the increased competition afforded to the market. However, it also presents new modelling, pricing and risk management challenges to both researchers and practitioners involved in the energy sector.

Electricity is a flow commodity, which means that it has to be delivered over time. It differs from other commodities in that it is characterized by non-storability and inelasticity of demand (Shiryaev et al., 2006, p. 239). Electricity also has specific price drivers such as weather conditions and the intensity of business and everyday activities. All in all, this results to it having price features seldomly seen in other commodity markets such as complex seasonal patterns, periods of huge volatility and extreme price spikes (Weron, 2014).

Due to the non-storability property of electricity, the only way to financially hedge against price uncertainties in the future is through the purchase of futures contracts, which promises the delivery of electricity at a predetermined price at a specified time in the future. However, the futures contracts available in the market today only offer certain specific delivery periods than sometimes desired. Thus, there is a need for an over-the-counter (OTC) market, where one can trade contracts with non-standard delivery periods. This is where the *Hourly Price Forward Curve (HPFC)* comes into play. A forward price curve is the price of a commodity, with delivery period in the future, as seen today. Hence, the HPFC is a forward curve with an hourly granularity, which can be used to value any forward position regardless if the delivery period is standard or non-standard.

Another consequence of its non-storability is that the standard forward pricing method cannot be applied and hence there is a need for an alternative pricing method. The solution is the HPFC and its construction can generally be decomposed into two steps. The first step is the estimation of the seasonal *shape vector*. The shape vector is a vector of hourly weights that reflects all seasonalities; intra-daily, intra-weekly and intra-yearly. Moreover, it also reflects weekends, holidays and bridge days. All forward-looking information is hence collected in the shape vector. The second step calibrates the estimated shape vector to prevailing forward prices by including arbitrage free conditions.

This thesis focuses on the first step of the aforementioned methodology, i.e. the estimation of the shape vector. Furthermore, this study is intended to examine the possibilities of estimating the shape vector using a parametric approach with external inputs.

1.1 Purpose and Research Questions

The purpose of this thesis is to investigate whether or not we can employ a parametric model to estimate the shape vector. Additionally, this thesis will examine the possibilities of including external regressors in the aforementioned parametric model. This study focuses exclusively on modelling the shape vector concerned with the month-to-hour ratios, i.e. hourly weights within each month. Thus, the aim of this thesis can be summarized as the following research questions

(i) Can we employ a parametric model to estimate the month-to-hour ratios?

(ii) Can we include external regressors in the parametric approach?

1.2 Related Articles

Several previous papers have investigated different HPFC construction methodologies.

In his paper, (Green, 2014) constructs an HPFC with hydrology dependence. The intra-daily and intra-weekly seasonalities are estimated using a feed-forward artificial neural network, which is trained on a historical data set of hourly spot prices from the Nord Pool market and weekly measurements of the Nordic hydrological balance. Moreover, the intra-yearly seasonality is estimated using historical futures prices from Nasdaq OMX. The shape vector consisting of all the estimated seasonal patterns is thereafter calibrated to an arbitrage free HPFC. Three different hydrological scenarios are defined; normal, wet and dry. It is found that a wet scenario increases the spread between peak and off-peak prices, while a dry scenario does the opposite, which is in line with historical data and the expected traits of hydro power generation.

In another study, (Kaffe, 2011) constructs an HPFC using linear programming. Using linear programming, the estimation and calibration of the shape vector is conducted simultaneously. The shape vector is estimated using two different regression techniques: Least Absolute Deviation (LAD) and LAD-LASSO, where LAD-LASSO is a combination of LAD and LASSO. The author argues that LAD regression is more robust against outliers than Ordinary Least Squares (OLS) regression. The calibration is then done using arbitrage free constraints and additional constraints in the linear programming set up. Kaffe trains the model on a historical data set of hourly spot prices, weather data, and futures prices, which is collected from Germany. It is concluded that the models illustrate reasonable results. Moreover, the addition of LASSO resulted in a general improvement in the model, where the HPFC became narrower, without extremities.

In a KYOS analysis report, (Beolet, de Jong, and Enev, 2014) study the effect of renewable power production on the intra-daily seasonalities. The study is conducted using data from Germany. The authors use regression to estimate the sensitivity of spot prices to wind and solar power production. Parameters are estimated for every hour of the day and for predefined day types using dummy variables. Using this regression technique, it is shown that a 1% rise in wind generation tends to provoke a 1-5% fall in the spot price. The dependency of spot prices on solar production is comparable to wind. Finally, when incorporating the renewable regression technique in the HPFC construction, the authors observe an overall improvement in fit. On average, it results in creating higher prices in the early evening and lower prices in the middle of the day.

1.3 Contribution Statement

The research regarding HPFC construction is extensive. However, the research on how to model the HPFC using multiple external inputs is sparse. Furthermore, previous papers have focused on the calibration procedure. Thus, the main contribution of this paper is that we will exclusively focus on modelling the shape vector and investigate the possibilities of including multiple dependencies when doing so.

1.4 Thesis Outline

This thesis is organized as follows. Chapter 2 briefly discusses and overviews the Nordic electricity market. Chapter 3 describes the HPFC and the methodology behind its construction. Chapter 4 overviews the data sets and the data pre-processing. In Chapter 5, the theoretical foundation behind the modelling approach is given. The theory behind the model evaluation process is also given in this chapter. Chapter 6 discusses the model design and implementation using the previously described theory. Furthermore, this segment discusses our approach regarding model selection. The results will be presented in Chapter 7. Finally, we state our conclusions and the results are evaluated and discussed in the following chapter. Chapter 9 outlines some suggestions for further work.

Chapter 2

Nordic Electricity Market

The recent deregulation has opened up the Nord Pool electricity market to free competition. The purpose of exposing the production and sale of electricity to free competition is to increase the consumers' freedom of choice and to create a more efficient market with an increased power exchange between countries and an increased security supply. Furthermore, the integrated market between muliple countries results in a more diversified market (Nord Pool Power Market, 2019).

A central feature of a liberalized electricity market is the existence of a power exchange and a market for financial instruments. A power exchange covers the physical trading of electricity, whereas the purely financial instruments are traded either bilaterally at the OTC-market or through a financial market such as Nasdaq OMX.

The Scandinavian deregulation led to the establishment of the Nordic electricity power exchange, Nord Pool. Other power exchanges in Europe include UKPX in the United Kingdom, EEX in Germany and OMEL in Spain. This paper exclusively focuses on the Nord Pool electricity market.

2.1 Nord Pool and the Physical Market

Nord Pool was established in the 1990s, initially covering Norway and Sweden. Today Nord Pool covers the four Scandinavian countries and three Baltic countries; Sweden, Norway, Denmark, Finland, Estonia, Latvia and Lithuania. Nord Pool provides two different markets for the physical trading of electricity; Elspot and Elbas. Elspot is the day-ahead market where power is traded a day in advance for every hour the following day. The day-ahead prices are determined through an auction procedure where buyers and sellers submit bid and ask prices. Elbas is however the intra-day market where power is traded continuously up to one hour before delivery. The intra-day market functions as a balancing market to Elspot (Energiinspektionen, 2017).

The Nord Pool market is currently divided into 15 bidding regions and they are illustrated in Figure 2.1. The price in each region, called the area price, might differ due to the fact that available transmission capacity may vary and congest the flow of power between the various bidding areas. In contrast to the area price, the system price is the Nordic reference price, which is calculated as the equilibrium between the aggregated supply and demand curves for all the bidding areas. Thus, the system price is not an average of the regional prices. The system price is calculated assuming no congestion and no transaction costs between the bidding areas and is thereby a fictive price. In this report, we are only studying the system price (Nord Pool Bidding Areas, 2019).

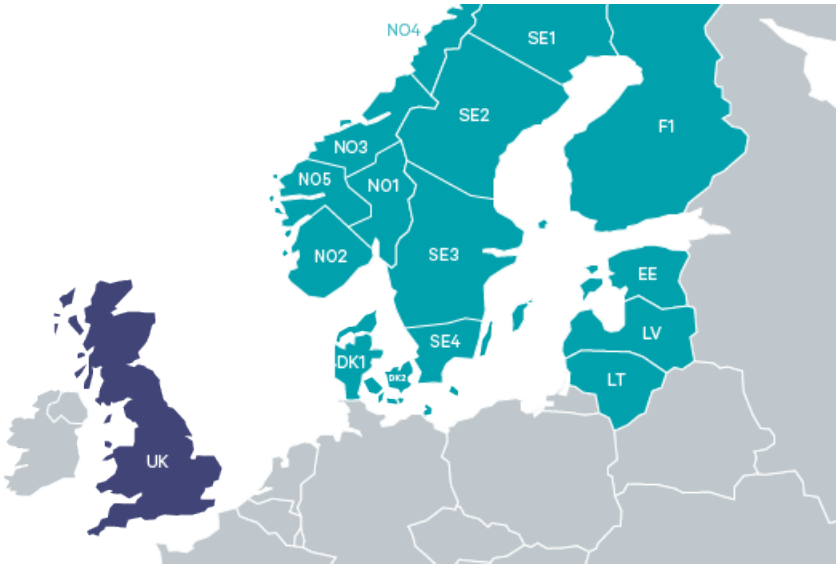


Figure 2.1: Nord Pool bidding areas.

2.2 Financial Markets

Financial instruments covering the electricity market are traded either at Nasdaq OMX or at the OTC-market. These financial contracts are primarily used for hedging against price uncertainties and other risk management purposes due to spot prices' volatile behaviour.

The product offering at Nasdaq OMX includes futures, deferred settlement futures (DS Futures), options and Electricity Price Area Differentials (EPAD) contracts. They all have the Nord Pool system price as their reference price. There is no physical delivery of electricity occurring here, only cash settlements. Moreover, these power derivatives can have different delivery periods; daily, weekly, monthly, quarterly and yearly. The delivery periods for the different products at Nasdaq OMX are shown in Table 2.1. The specifications of the power derivatives are briefly described in the next paragraph (Nasdaq OMX, 2019).

A power futures contract is an agreement between two parties to trade electricity during a predetermined delivery period, at a price locked in today. They are settled daily against the Nord Pool system price. Thus, if the futures are sold at a price exceeding the system price, the seller is compensated by the buyer equivalent to the difference, and vice versa. In contrast to traditional futures contracts, deferred settlement futures are only settled at the delivery period and not daily. Both futures and deferred settlement futures can be used to hedge against the system price. In order to hedge against price area risk caused by constraints in the transmission grid, it is possible to trade EPAD contracts, which reflect the difference between the area price of a bidding region and the Nord Pool system price. These are offered as both futures and DS futures. Finally, electricity options are offered, which gives the holder the right to trade the underlying asset at a fixed price on a prespecified date (Nasdaq OMX, 2019).

If an agent is interested in trading a non-standard financial contract, it has to be priced and it is done so using the HPFC. These forward contracts are not traded at a centralized exchange such as Nasdaq OMX. Instead they are traded at the OTC-market.

Product Offerings	**Delivery Periods**
Futures	
Nordic Electricity	Daily, Weekly, Monthly, Quarterly, Yearly
EPADS	Weekly, Monthly, Quarterly, Yearly
DS Futures	
Nordic Electricity	Monthly, Quarterly, Yearly
EPADS	Monthly, Quarterly, Yearly
Options	
Futures	Monthly, Quarterly, Yearly
DS Futures	Quarterly, Yearly
Minimum Contract Size	1 MW
Currency	EUR

Table 2.1: Power derivatives available at Nasdaq OMX and their respective delivery periods.

Chapter 3

Hourly Price Forward Curves

3.1 Definition

The traditional forward pricing method for commmodities includes pricing the future contract F_0 by discounting the current spot price S_0 and storage costs U with interest rate r and time to maturity T. The formula is presented in (3.1)

$$F_0 = (S_0 + U)e^{-rt} \tag{3.1}$$

However, this method is not applicable for non-storable commodities and thereby an alternative pricing method is needed (Hildmann, Herzog, et al., 2011).

The pricing method used by market participants is the HPFC. It essentially gives the price of electricity with delivery period in the future, at an hourly resolution, given market information available today. Thus it contains inherent seasonality characteristics displayed by the spot price and long-term market expectations. It is not only used to price non-standard products at the OTC-market, but also serves as an essential input when evaluating risks and conducting profitability calculations of power plants (Hildmann, Caro, et al., 2017).

3.2 Construction of the HPFC

Constructing an HPFC generally consists of the following two steps

(i) Estimation of the shape vector

(ii) Price calibration of the estimated shape vector to forward prices

3.2.1 Shape Estimation

A requirement of the HPFC is that it must capture the intra-yearly, intra-weekly and intra-daily seasonalities. Furthermore, it must also account for patterns occurring due to public holidays, weekends and bridge days (Hildmann, Caro, et al., 2017). All of these different periodicities are reflected in the estimated shape vector containing hourly weights normalized for every year, which we choose to call the year-to-hour ratios in this thesis.

There are three modelling approaches when estimating the shape vector

1. The **statistical** approach, which is based on statistical models on historical data
2. The **fundamental** approach, which is based on the supply and demand curve
3. The **hybrid** approach, which combines the two above-mentioned models

This thesis takes on a statistical approach and it does so taking both seasonality and external factors into account. Moreover, this paper is only concerned with modelling the month-to-hour ratios, i.e. a shape vector with hourly weights normalized for every month. The mean of the weights for each month must therefore sum to unity. Moreover, the shape vector is only constructed using the Nord Pool system price.

3.2.2 Price Calibration

Another requirement of the HPFC is that it must be arbitrage free, which is achieved through the price calibration procedure where the expected value of the HPFC must equal to the corresponding traded futures contract. This is generally achieved through solving an optimization problem with arbitrage free constraints using products available at the market (Sviland Saetherö, 2017). However, this report does not cover this step of the HPFC construction; only the first step.

3.3 Validation of the HPFC

The validation and evaluation of the HPFC is not unambiguous. There is no established consensus among market participants regarding this topic. Furthermore, different energy companies have different HPFCs as a result of differing construction methods. Generally, this information is confidential, which makes it difficult to compare different HPFC methodologies. Moreover, it is important to note the difference between a forward curve and a price forecast. The forward curve represents the long-term market expectations based on market information available today. However, it is not a prediction of future spot prices and it is therefore not expected to replicate its volatile behaviour, only its overall pattern characteristics. In this thesis, we have used different validation metrics to evaluate our constructed shape vector, which is presented in Section 5.5. However, it is crucial that the shape vector also illustrates specific traits and fulfills certain criteria. Recommended quality measurements (Hildmann, Caro, et al., 2017), applicable to our case, include

(i) *Visible seasonal patterns:* The shape must contain the right level of variation between the hours of the day (intra-daily pattern). The intra-daily pattern must capture the peak and off-peak differences. Furtermore, the shape must also deviate between weekdays and weekends (intra-weekly), where the demand is mainly related to the working hours during weekdays, whereas the pattern is more sporadic during weekends. Finally, it must account for intra-yearly seasonalities, where the higher demand during the winter in comparison to the summer, must be reflected in the shape.

(ii) *Robust against outliers and overfitting:* The shape should be robust against the extreme and volatile behaviour of spot prices. It should also be robust against overfitting unique events. For an instance, if a January month shows higher prices than usual due to non-standard external effects, the shape should be robust against this in its predictive performance.

In order to make sure the shape contains the specific aforementioned traits, a graphical inspection is needed as a complement to the validation metrics.

Chapter 4

Data

This chapter covers the data used in this thesis and the necessary data pre-processing before initiating the model implementations. The data set is provided by E.ON and Nord Pool, where the data from Nord Pool is publicly available through their website. This segment is divided into two parts, where the first one covers the external inputs (fundamentals) that are fed to the models, and the second one describes the target variable (month-to-hour ratios) being modelled. The data set covers the Nord Pool market and it ranges from January 1, 2008 to February 10, 2019. Moreover, the training set ranges from January 1, 2008 to December 31, 2016, and the test set was set between January 1, 2017 and February 10, 2019.

4.1 Fundamentals

The following fundamentals are investigated: hydrological balance, consumption, temperature deviation, cross-border flow and installed wind capacity. The chosen variables is a result of a compromise between data availability and discussions with analysts at E.ON.

4.1.1 Hydrological Balance

The hydrological balance is defined as the deviation from the normal state of hydrological reservoirs, which in turn is a measure of the amount of potential resources for hydro power generation. Since more than 50% of the Nordic power generation stems from hydro sources, this is an important price driver and thus affects the behaviour of the shape vector.

Spot prices respond differently depending on the change in hydrological balance. In cases of a hydrological oversupply, production can be difficult to control during low-demand hours. This may result in a severe drop in spot prices. However, for hours when consumption is high, it is easier to manage the generated supply from hydro and prices exhibit normal levels. The combined effect is that prices generally exhibit a larger spread between peak and off-peak hours due to difficulties of controlling the flow. On the contrary, a scenario of hydrological undersupply will result in tighter price spreads (Green, 2014). These types of characteristics combined with the fact that hydro supplies a large share of electricity in the Nord Pool region, makes it natural for us to examine its possibilities to model the shape vector.

We were given weekly measurements of the hydrological balance in GWh, which we linearly interpolated to daily measurements with every hour within the same day having the same value. Figure 4.1 depicts the Nord Pool hydrological balance for the entire data set.

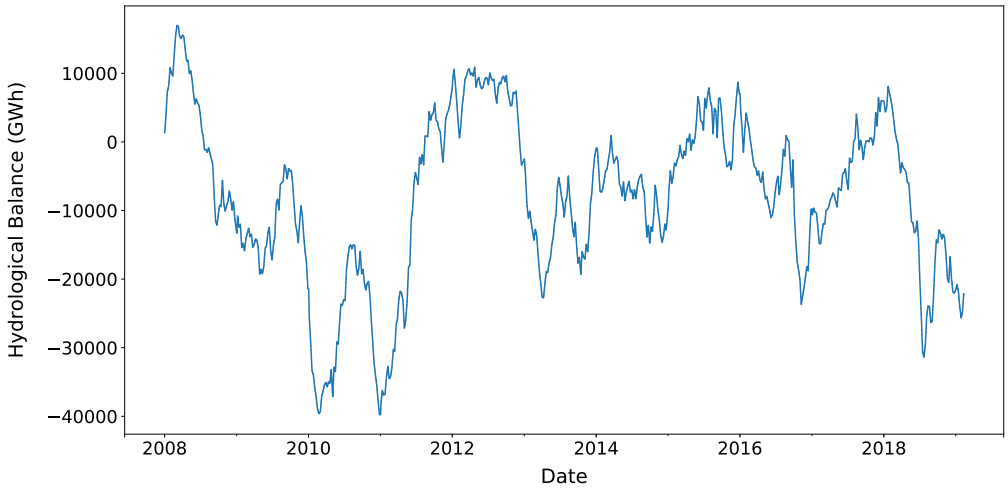


Figure 4.1: The Nordic hydrological balance for the entire data set.

4.1.2 Consumption

The second investigated fundamental is the total consumption level of the Nord Pool region.

The electricity market is characterized by an inelastic demand compared to the supply curve. This means that high demand can result in extreme price peaks since the buyers are not price sensitive. Since consumption reflects the demand it should contain information, not only regarding the actual demand itself, but also regarding the aforementioned price peaks, and thereby being useful when modelling the shape vector.

Another unique characteristic for electricity markets, which makes consumption interesting, is that due to electricity's non-storability, the production and consumption must be in constant balance (Kaminski, 2013, p. 755). Here the parties are obliged to trade, unlike in other financial markets where you trade if you agree on the price. This adds a dimension to the relationship between electricity prices and consumption, which is difficult to predict.

Hourly measurements of the Nord Pool consumption in GWh were given and is displayed in Figure 4.2. Since the consumption varies within a day, week and year, it contains important information about these seasonalities. In Figure 4.2 the intra-yearly seasonality is apparent, where we observe higher consumption levels in the winter compared to the summer, which can be attributed to the cold winters in the Nordics.

4.1.3 Temperature Deviation

To examine the relationship with temperature, we have included hourly data containing the deviation from the normal temperature. This data is used to explain some of the non-seasonal weather dependencies in the month-to-hour ratios. One limitation arising from this data is that the average temperature of Sweden and Norway is used, which is not an entirely accurate representation of the temperature in the entire Nord Pool region. However, a Nord Pool reference temperature was not available and thereby the data at hand was investigated instead.

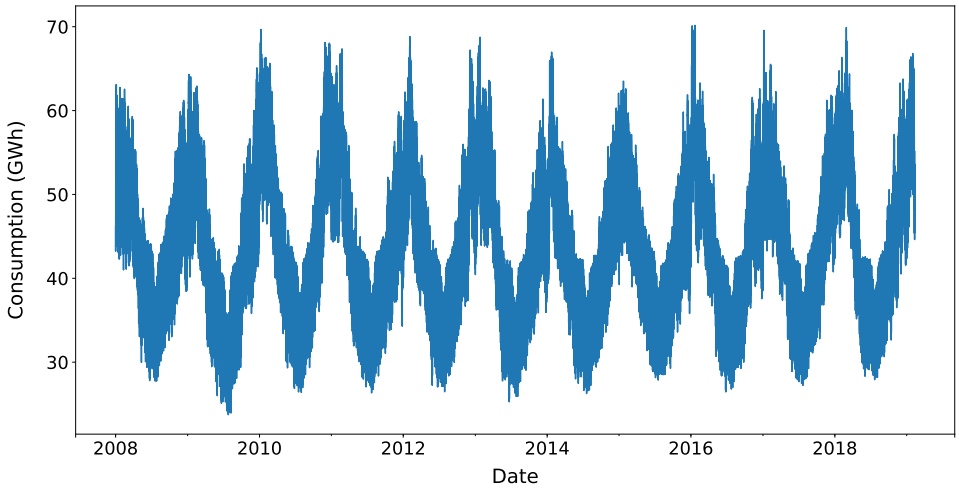


Figure 4.2: The total consumption in the Nord Pool region for the entire data set

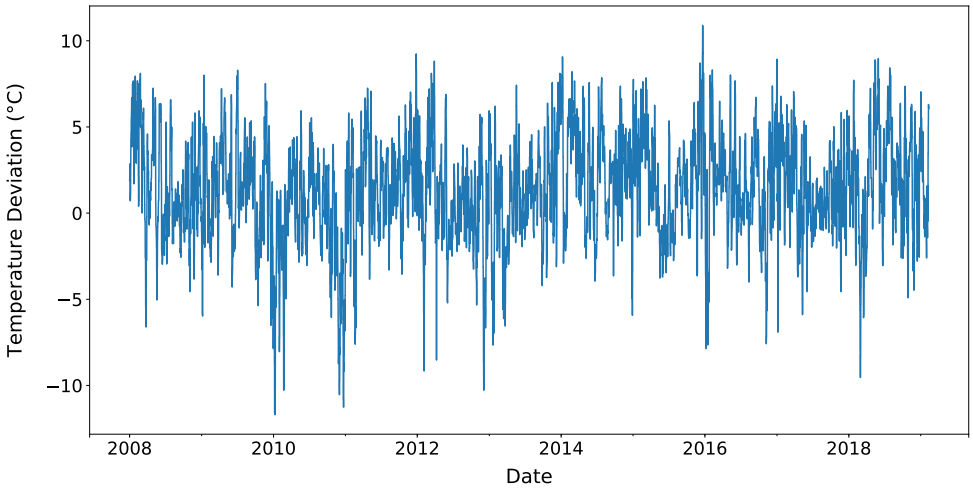


Figure 4.3: The temperature deviation for Sweden and Norway for the entire data set.

4.1.4 Cross-Border Flow

The net exchange of the flow into and out of the Nord Pool region is also studied to determine the import/export effects on the spot prices and thus the shape vector. The data was given with an hourly resolution.

In 2014 the European Commission set a goal for the member states to have an interconnection of at least 10% of their power production by 2020 (The European Commission, 2017). One of the purposes with this is to reduce price differences between member states. The idea is that a market can control its supply by trading with other markets. E.g. if the Nord Pool region has an oversupply, it can sell some of it to an interconnected system such as the German or the Dutch. Therefore, quantities traded in and out of Nord Pool may contain relevant information about the price levels when modelling the shape vector.

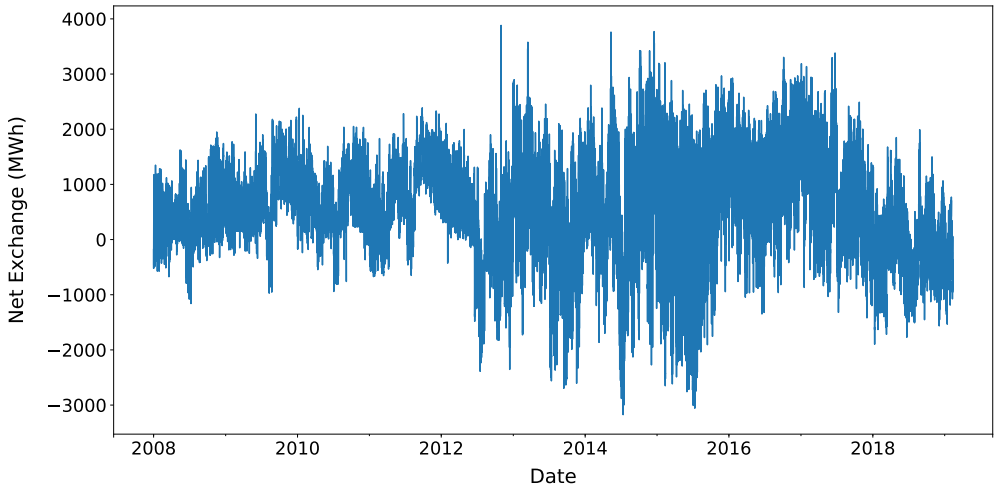


Figure 4.4: The net exchange of the flow in the Nord Pool region for the entire data set.

4.1.5 Installed Wind Capacity

The shape vector should also extract necessary information regarding the relationship between electricity spot prices and the expansion of renewables. As mentioned in Section 1.2, (Beolet, de Jong, and Enev, 2014) observed an overall improvement when including renewables to construct their HPFC. Studying renewables is thus of uttermost importance considering the ongoing expansion of renewables in the Nord Pool region. In the Nordics, wind power expansion has progressed strongly. Wind energy has unique characteristics; its supply is significantly determined by meteorological conditions, it has low marginal costs in comparison to other energy sources and it has been shown to increase the volatility of prices (Brancucci Martinez-Anido, Brinkman, and Hodge, 2016).

In this thesis we have included installed wind capacity in Nord Pool as an explanatory variable, which is illustrated in Figure 4.5. It is of interest to study the wind power production, but due to lack of data and difficulties extrapolating and conducting accurate predictions, it has been neglected from this study. It should however be acknowledged that it is quite difficult to model a stationary process with non-stationary data, but since this was the only available data we decided to include it.

4.2 Target Variable

As previously mentioned, this study has been restricted to modelling the month-to-hour ratios. This has been constructed by dividing the price at each hour t with its monthly average. The month-to-hour ratios and the hourly prices are denoted by y_t and p_t respectively, where the construction of the ratios for a month with N hours is as follows

$$y_t = \frac{p_t}{\frac{1}{N}\sum_{i=1}^{N} p_i} \tag{4.1}$$

and therefore they have the property that their average sums to one over the month.

The month-to-hour ratios are illustrated in Figure 4.6, where the training and test sets are differentiated as well.

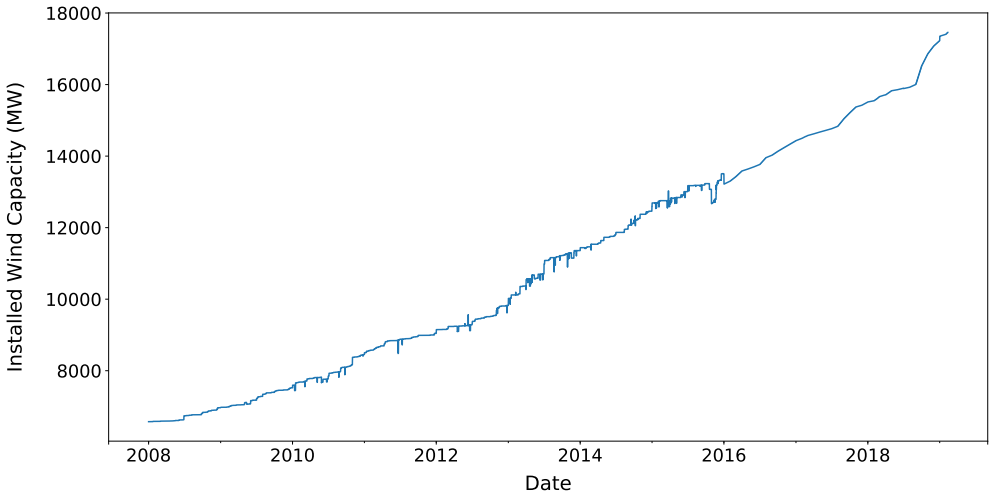


Figure 4.5: Total installed wind capacity for the entire dataset.

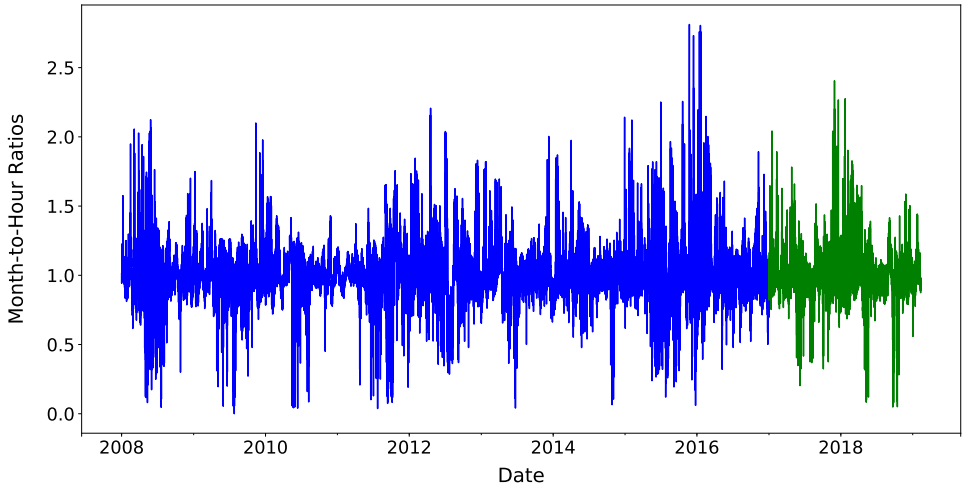


Figure 4.6: The month-to-hour ratios for the training set (blue) and the test set (green).

4.2.1 Electricity Spot Price Characteristics

The hourly electricity demand and hence spot price can be explained by the following key factors (Burger, Graeber, and Schindlmayr, 2014, p. 304)

(i) *Season:* Patterns of electricity use and hence spot price show typical seasonal behaviour, which is simply due to varying duration of day-light, seasonal production patterns of some industries and demand for heating and cooling.

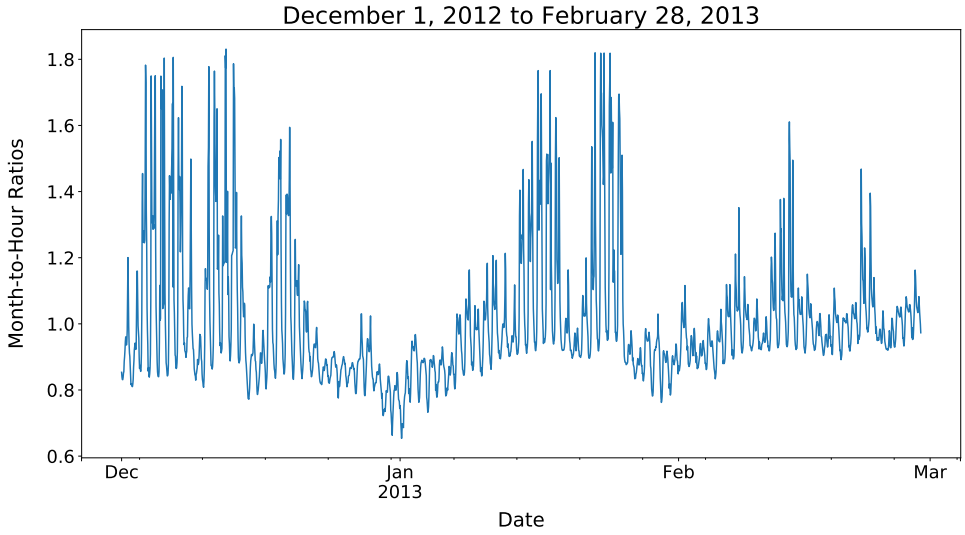
(ii) *Day of the week:* Spot prices exhibit strong differences between different weekdays and weekend days.

(iii) *Irregular sectoral non-work or limited-activity days:* These include public holidays and bridge days. The demand of industry and services are often lower on these days, while household demand might

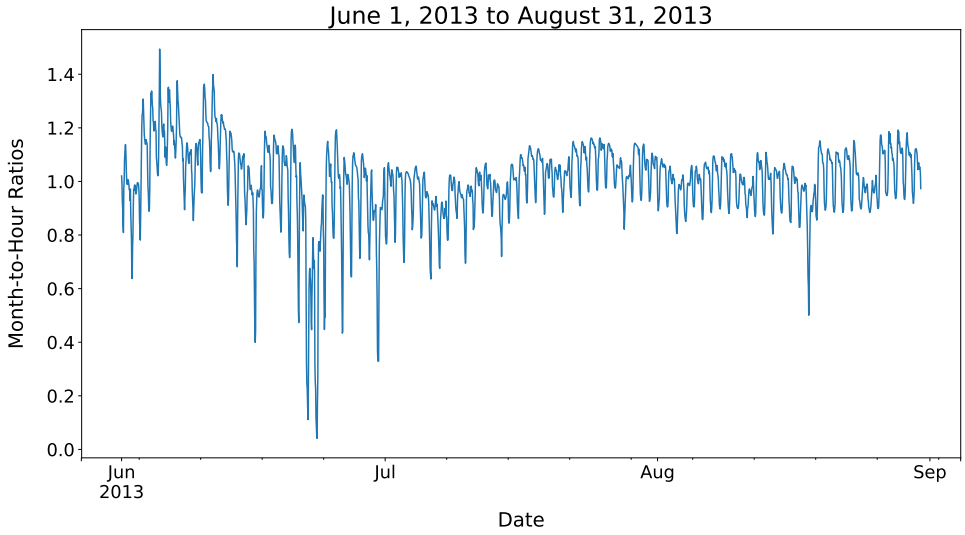
increase.

(iv) *Hour of the day:* Spot prices follows typical intra-daily patterns, which differ between different day types. Regarding the intra-daily patterns, spot prices generally display a pronounced morning peak and evening peak due to higher demand.

The intra-yearly pattern is illustrated in Figure 4.7, where the month-to-hour ratios from December 2012 to February 2013 and from June 2013 to August 2013 have been plotted. Note that the ratios are more spiky during the winter than the summer. This is due to generally higher prices during winter months and larger price spreads between low-demand and high-demand hours (off-peak and peak hours).



(a) Month-to-hour ratios from December 1, 2012, to February 28, 2013.

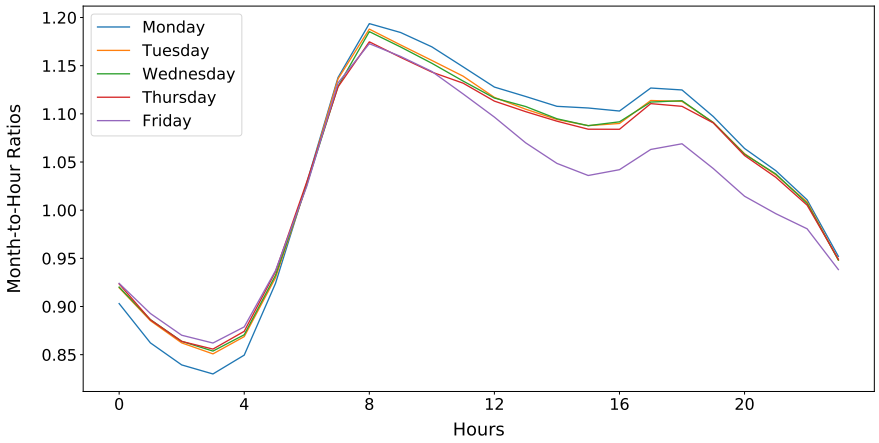


(b) Month-to-hour ratios from June 1, 2013, to August 31, 2013.

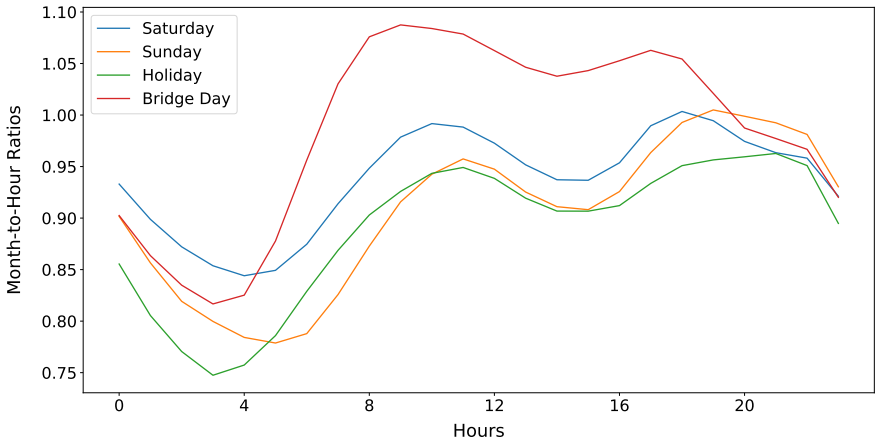
Figure 4.7: Examples of month-to-hour ratios from parts of the training period.

Moving forward, we decided to form clusters of days (day types) with similar or unique characteristics. The following day types were created using dummy variables: Monday-Thursday, Friday, Saturday, Sunday, Public Holidays, and Bridge Days. This was decided by analyzing the plots in Figure 4.8 depicting the intra-daily seasonalities. Figure 4.8a and Figure 4.8b show the average month-to-hour ratio of every

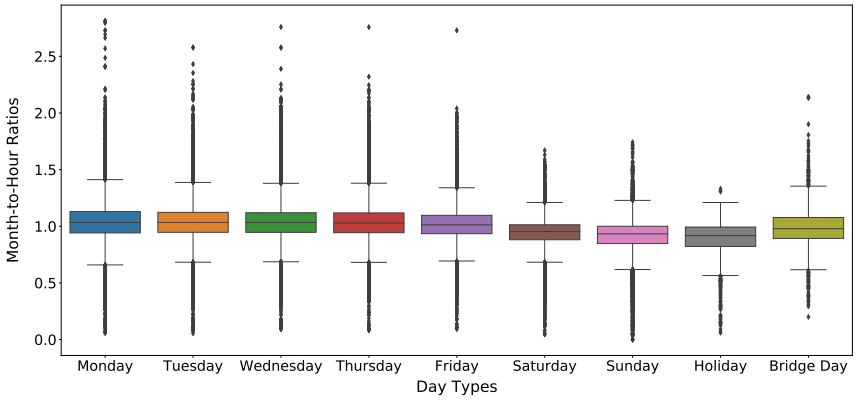
hour of every day type, before categorizing them. Figure 4.8c illustrates the variation of the ratios for each day type using a box plot.



(a) Comparison of the average month-to-hour ratios per hour between working days over the training set.



(b) Comparison of the average month-to-hour ratios per hour between non-working days over the training set.



(c) Box plot of the month-to-hour ratios per day type over the training set.

Figure 4.8: Average month-to-hour ratios per day type over the training period.

We immediately observed a difference between working days and non-working days, leading us to divide them into two categories. Moreover, when comparing the working days with each other, Friday was the only day having ratios significantly differing from the others, motivating the choice of Monday-Thursday and Friday as two separate categories. However, moving forward with the non-working days, they were more challenging to find similarities between as seen in Figure 4.8b, resulting in one category for each day. Especially public holidays and bridge days are challenging to model due to the limited amount of historical data. This decision was also taken after discussions with analysts at E.ON.

The spot price data was provided by Nord Pool and E.ON. The most recent data was collected from Nord Pool, whereas the data prior to 2013 was provided by E.ON. The data contains hourly Nord Pool system prices. However, due to summertime adjustments in Europe, the data set contains a missing value the last Sunday in March (when clocks are advanced by one hour), and duplicate values the last Sunday in October (when clocks are set back by one hour). We convert the data to a 24 hour format by replacing the missing value with its previous value and by removing the duplicate.

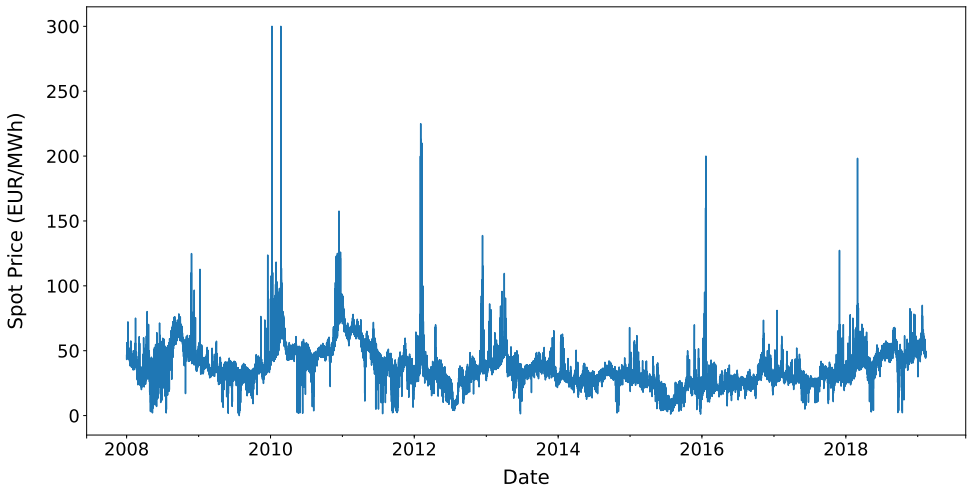
Note that this study does not aim to model the volatile and extreme behaviour characterized by spot prices, but to capture the overall seasonal patterns. For this reason, there is a need to adjust this data for outliers, which is discussed in the next section.

4.2.2 Outliers

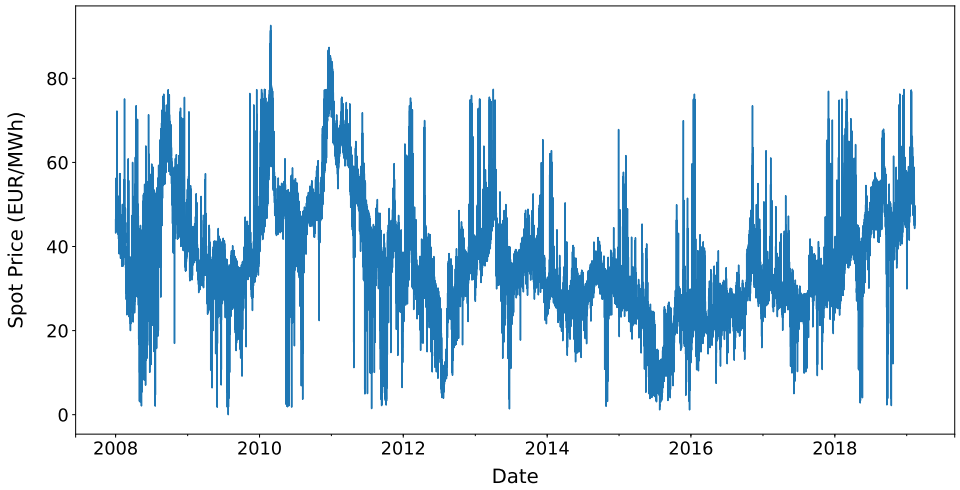
Regarding missing values, error values, and outliers among the fundamentals, these were treated with linear interpolation. However, when adjusting the target variable for outliers, a different approach was taken. Figure 4.9a illustrates the spot prices for the entire data set. This data needs to be adjusted for the spikes and after experimenting with different percentiles we decided to regard the spot prices above the 99th quantile for the entire dataset as outliers. These were replaced by its 168 hour moving average to smooth out any possible effects due to volatility clustering characterized by spot prices. The month-to-hour ratios were thereafter constructed using these outlier-adjusted prices illustrated in Figure 4.9b.

4.2.3 Data Pre-Processing

Figure 4.10 summarizes the data pre-processing methodology before proceeding to the modelling stage.



(a) Nord Pool system prices.



(b) Outlier-adjusted Nord Pool system prices.

Figure 4.9: Nord Pool system prices before and after outlier-adjustment for the entire data set.

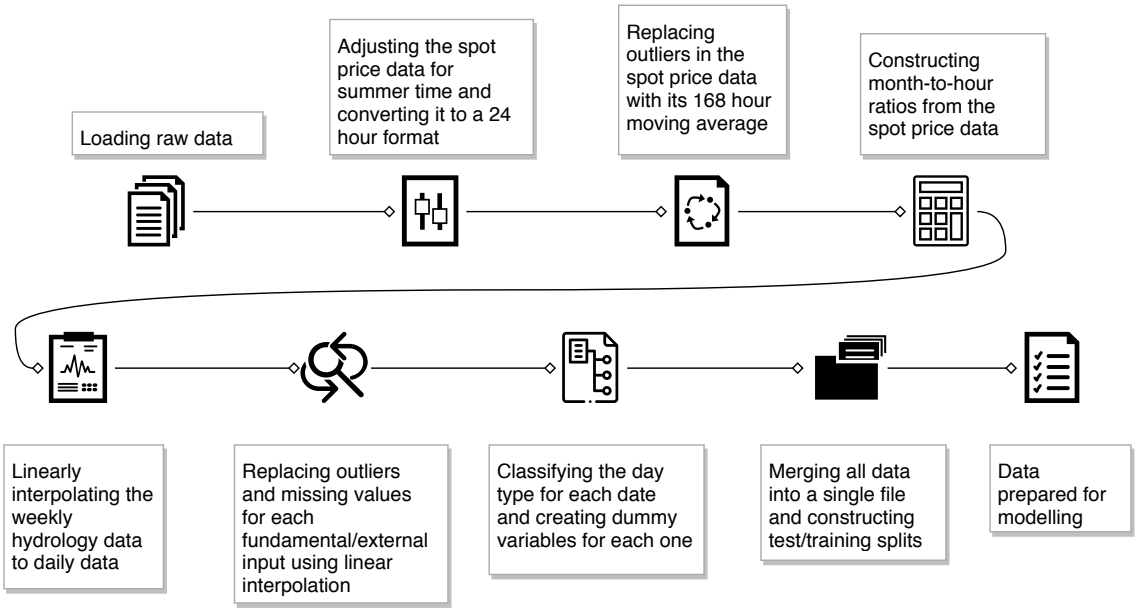


Figure 4.10: The main steps in the data pre-processing.

Chapter 5

Theory

This chapter will present the theoretical concepts used in the modelling approach.

5.1 Linear Regression

A linear regression model in its simplest form assumes that the dependent variable y can be explained by a linear combination of the input variables,

$$y(\mathbf{x}, \mathbf{w}) = w_0 + w_1 x_1 + ... + w_N x_N \tag{5.1}$$

or with matrix notation,

$$y = \mathbf{w^T x} \tag{5.2}$$

where the input variables are $\mathbf{x} = (x_1, ..., x_N)^T$ and the parameters are $\mathbf{w} = (w_0, ..., w_N)^T$.

When considering observations of the dependent variable y_i an error term ϵ_i is added.

$$y_i = \mathbf{w^T x_i} + \epsilon_i \tag{5.3}$$

The error terms are assumed to have zero mean, have the same variance σ^2 and to be pairwise independent.

This model is a linear function of both the parameters and the input variables. However, it may be extended by adding linear combinations of non-linear functions of the input variables, called basis functions $\phi(\mathbf{x})$,

$$y(\mathbf{x}, \mathbf{w}) = \sum_{j=0}^{N} w_j \phi_j(x) = \mathbf{w^T \boldsymbol{\phi}(x)} \tag{5.4}$$

where $\mathbf{w} = (w_0, ..., w_N)^T$ and $\boldsymbol{\phi} = (\phi_0, ..., \phi_N)^T$ (Bishop, 2006, p. 138-139).

5.1.1 Ordinary Least Squares

Ordinary Least Squares (OLS) solves the linear regression problem by estimating the parameters that minimize the sum of squares of the residuals, introduced as the error function $Q(\mathbf{w})$.

$$Q(\mathbf{w}) = \left(\mathbf{y} - \mathbf{w^T \boldsymbol{\phi}(x)}\right)^T \left(\mathbf{y} - \mathbf{w^T \boldsymbol{\phi}(x)}\right) \tag{5.5}$$

The parameters minimizing $Q(\mathbf{w})$ are calculated using the normal equations

$$\boldsymbol{\phi}(\mathbf{x})^T \boldsymbol{\phi}(\mathbf{x}) \hat{\mathbf{w}} = \boldsymbol{\phi}(\mathbf{x})^T \mathbf{y} \tag{5.6}$$

giving the solution for the estimated parameters $\hat{\mathbf{w}}$.

$$\hat{\mathbf{w}} = \left(\boldsymbol{\phi}(\mathbf{x})^T \boldsymbol{\phi}(\mathbf{x})\right)^{-1} \boldsymbol{\phi}(\mathbf{x})^T \mathbf{y} \tag{5.7}$$

When applying OLS, some assumptions are required to be fulfilled for the solution to be considered BLUE (Best Linear Unbiased Estimator). The assumptions are as follows (Rawlings, Pantula, and Dickey, 1998, p. 77).

(i) Linearity of the relationship between dependent and independent variables.

(ii) The error terms are homoscedastic, independent and identically distributed.

(iii) No perfect collinearity between the independent variables.

5.2 Basis Functions

5.2.1 Polynomial Basis Functions

The polynomial basis function calculates the power of the variable x.

$$\phi_j(x) = x^j \tag{5.8}$$

The polynomial basis can then be constructed as follows

$$\boldsymbol{\phi}_{poly}(x) = \begin{bmatrix} \phi_0(x) \\ \phi_1(x) \\ \phi_2(x) \\ \vdots \\ \phi_N(x) \end{bmatrix} = \begin{bmatrix} 1 \\ x \\ x^2 \\ \vdots \\ x^N \end{bmatrix} \tag{5.9}$$

5.2.2 Fourier Basis Functions

Fourier basis functions are defined as sine and cosine functions for a certain frequency ω.

$$\boldsymbol{\phi}(t) = \begin{bmatrix} \sin(\omega t) \\ \cos(\omega t) \end{bmatrix} \tag{5.10}$$

When including Fourier basis functions in a linear regression model it is to capture periodicities in the data. The basis functions is formed with increasing frequencies as follows

$$\boldsymbol{\phi}_{Fourier}(t) = \begin{bmatrix} \boldsymbol{\phi}_0(t) \\ \boldsymbol{\phi}_1(t) \\ \boldsymbol{\phi}_2(t) \\ \vdots \\ \boldsymbol{\phi}_N(t) \end{bmatrix} = \begin{bmatrix} 1 \\ \sin(\omega t) \\ \cos(\omega t) \\ \sin(2\omega t) \\ \cos(2\omega t) \\ \vdots \\ \sin(N\omega t) \\ \cos(N\omega t) \end{bmatrix} \tag{5.11}$$

resulting in an orthogonal base (Ramsay and Silverman, 2005, p. 45).

5.3 Regularization

When conducting regression analysis there is always a risk of overfitting, i.e. that the model explains the training data so well to an extent that it fails to predict additional data reliably. A solution to this

problem is to introduce a regularization term in the error function in addition to the data-dependent error, resulting in an error function defined as $Q(\mathbf{w})$

$$Q(\mathbf{w}) = Q_D(\mathbf{w}) + \lambda R_W(\mathbf{w}) \tag{5.12}$$

where Q_D is the data-dependent error (e.g. the sum of squares of the residuals), R_W is the regularization term and λ is the regularization coefficient. This coefficient determines the regularization term's importance in proportion to the data-dependent error (Bishop, 2006, p. 144-146).

5.3.1 Elastic Net

Elastic Net is a regularized regression technique with a regularization penalty combining the l_1-norm and the l_2-norm of the parameter vector $\mathbf{w}$

$$\frac{1}{2}(1-\alpha)\left\|\mathbf{w}\right\|_2^2 + \alpha\left\|\mathbf{w}\right\|_1 \tag{5.13}$$

where $\alpha \in [0,1]$ decides the weights of the norms. Combining the Elastic Net penalty with the OLS sum of squared of residuals gives the following error function for the regression model to minimize.

$$Q(\mathbf{w}) = \left(\mathbf{y} - \mathbf{w^T}\boldsymbol{\phi}(\mathbf{x})\right)^T \left(\mathbf{y} - \mathbf{w^T}\boldsymbol{\phi}(\mathbf{x})\right) + \lambda\left[\frac{1}{2}(1-\alpha)\left\|\mathbf{w}\right\|_2^2 + \alpha\left\|\mathbf{w}\right\|_1\right] \tag{5.14}$$

Elastic Net is essentially a compromise between the LASSO (Least Absolute Shrinkage and Selection Operator) and the Ridge regression techniques. These methods have different characteristics, which the Elastic Net tries to capture (Hastie, Tibshirani, and Wainwright, 2015, p. 56-58).

5.3.2 Ridge

Ridge is the special case of Elastic Net when $\alpha = 0$ and thus the penalty term is $\lambda\frac{1}{2}\left\|\mathbf{w}\right\|_2^2$. Since it is a quadratic regularizer it quickly shrinks large coefficients, removing the impact of highly correlated input variables. However, it rarely sets the coefficients to zero limiting the options of ruling out certain parameters (Hastie, Tibshirani, and Friedman, 2009, p. 61-68).

5.3.3 LASSO

LASSO is the special case of Elastic Net when $\alpha = 1$ and thus the penalty term is $\lambda\left\|\mathbf{w}\right\|_1$. Since LASSO penalizes on the absolute value it is a convenient tool for conducting variable selection, but not as efficient in shrinking sizeable coefficients (Hastie, Tibshirani, and Friedman, 2009, p. 68-69).

5.4 Time Series

5.4.1 Autoregressive model

The autoregressive model of order p, AR(p), is defined as

$$y_t + a_1 y_1 + ... + y_p a_p = e_t \tag{5.15}$$

where $a_p \neq 0$ and e_t is a zero-mean white noise error term. For the process to stay stationary it is necessary that $|a_p| < 1$ (Jakobsson, 2016, p. 68).

5.4.2 Kalman filter

The Kalman Filter uses observations of input and output of a system to construct the optimal linear estimators of the underlying unobservable state space. This is an iterative process and the estimators are updated following each new observation. The linear state space representation,

$$\mathbf{x}_{t+1} = \mathbf{F}\mathbf{x}_t + \mathbf{B}\mathbf{u}_t + \mathbf{w}_t \tag{5.16}$$

$$\mathbf{y}_t = \mathbf{H}\mathbf{x}_t + \mathbf{v}_t \tag{5.17}$$

describes the dynamics of the discrete time process that the Kalman Filter is applied to. $\mathbf{x}_t$ represents the hidden state, $\mathbf{y}_t$ is the measured output, $\mathbf{u}_t$ is a known input to the system, $\mathbf{v}_t \sim \mathcal{N}(0, \mathbf{R})$ and $\mathbf{w}_t \sim \mathcal{N}(0, \mathbf{Q})$ represents the Gaussian measurement and process noise respectively, all at time t. Moreover $\mathbf{F}$, $\mathbf{B}$ and $\mathbf{H}$ are known matrices describing the relations between the state, input and output.

The Kalman Filter assumes that the underlying process is Gaussian and thus the mean and the variance is what is required to conduct the state estimation. The algorithm itself can be split up into the steps prediction and update. The prediction relates to estimating the mean and the variance of the state in the next time period.

$$\hat{\mathbf{x}}_{t+1|t} = \mathbf{F}\hat{\mathbf{x}}_{t|t} + \mathbf{B}\mathbf{u}_t \tag{5.18}$$

$$V[\hat{\mathbf{x}}_{t+1}] = \mathbf{P}_{t+1|t} = \mathbf{F}\mathbf{P}_{t|t}\mathbf{F}^T + \mathbf{Q} \tag{5.19}$$

The update of the filter is described with the equations below. First the Kalman gain, equation (5.21), is constructed and thereafter used when updating the mean and the variance, deciding the influence of the observations on our new estimates of the parameters (Jakobsson, 2016, p. 290-294).

$$\mathbf{S}_{t|t-1} = \mathbf{H}\mathbf{P}_{t|t-1}\mathbf{H}^T + \mathbf{R} \tag{5.20}$$

$$\mathbf{K}_t = \mathbf{P}_{t|t-1}\mathbf{H}^T[\mathbf{S}_{t|t-1}]^{-1} \tag{5.21}$$

$$\hat{\mathbf{x}}_{t|t} = \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{H}\hat{\mathbf{x}}_{t|t-1}) \tag{5.22}$$

$$\mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t\mathbf{H})\mathbf{P}_{t|t-1} \tag{5.23}$$

$$\tilde{\mathbf{y}}_{t|t} = \mathbf{y}_t - \mathbf{H}\hat{\mathbf{x}}_{t|t} \tag{5.24}$$

5.5 Model Evaluation

5.5.1 Root-Mean-Squared Error

The root-mean-squared error (RMSE) is frequently used when evaluating predictive models. It summarizes the squared residuals, calculates the average and finally takes the root of the average (Bishop, 2006, p.6). It is introduced below as Q_{RMSE}.

$$Q_{RMSE}(\mathbf{w}) = \sqrt{\sum_{i=1}^{N} \big(y_i - y(x_i, \mathbf{w})\big)^2} \tag{5.25}$$

5.5.2 Mean-Absolute Error

The mean-absolute error (MAE) is the mean of the absolute error between two variables. The formula for Q_{MAE} is stated below.

$$Q_{MAE}(\mathbf{w}) = \sum_{i=1}^{N} |y_i - y(x_i, \mathbf{w}))| \tag{5.26}$$

5.5.3 R-Squared

The coefficient of determination, or R-Squared (R^2), is the fraction of the variance in the data that a model can predict. For predictive models, this is a very important measure and therefore included in the thesis (Rawlings, Pantula, and Dickey, 1998, p. 9).

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - y((x)_i, \mathbf{w}))^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2} \tag{5.27}$$

5.5.4 Price Prediction Errors

Although the HPFC does not aim to make an exact prediction of spot prices, it is interesting to compare the price errors of the models. To measure the performance in price prediction, the following two models were used (Sviland Saetherö, 2017). The first calculates the mean of the absolute difference between the estimated hourly prices and the realized prices, whereas the second calculates the mean squared error between the same two variables. The formulas are stated below

$$Aboslute\_Error = \frac{1}{M}\sum_{i=1}^{M}\frac{1}{N}\sum_{j=1}^{N}|MonthAveragePrice_i \times MonthToHourRatio_{i,j} - HourPrice_{i,j}| \tag{5.28}$$

$$Squared\_Error = \frac{1}{M}\sum_{i=1}^{M}\frac{1}{N}\sum_{j=1}^{N}(MonthAveragePrice_i \times MonthToHourRatio_{i,j} - HourPrice_{i,j})^2 \tag{5.29}$$

where j is the hour of the month i.

Chapter 6

Model Description and Implementation

When implementing the models, we have taken two different approaches. The first approach assumes a deterministic time-varying model, whereas the second one assumes a stochastic time-varying model.

The deterministic model was constructed using linear regression with polynomial and Fourier basis functions, using the fundamentals described in Chapter 4 as inputs. Furthermore, in this step, regularization techniques were used to avoid overfitting when increasing model complexity and as a model selection method. Finally, the stochastic model was implemented with variations of the Kalman filter estimating the parameters recursively.

6.1 Deterministic Time-Varying Model

The deterministic model includes polynomial bases to measure the dependencies with the fundamentals, Fourier bases to capture the different periodicities, and dummy variables to measure the effect of the predefined day types. The model was set up as follows

$$y_t = \mathbf{w}^{\mathbf{T}} \boldsymbol{\phi}(\mathbf{x}_t) + e_t \tag{6.1}$$

where $\boldsymbol{\phi}(\mathbf{x}_t)$ was constructed as follows

$$\boldsymbol{\phi}(x_t) = \begin{bmatrix} \boldsymbol{\phi}_{poly}(\mathbf{x}_t) \\ \boldsymbol{\phi}_{Fourier}(t) \\ \boldsymbol{\psi}_{dummy}(t) \end{bmatrix} \tag{6.2}$$

The function $\boldsymbol{\psi}_{dummy}(t)$ represents a vector of indicator functions, one for each day type defined in Chapter 4.2.1.

$$\boldsymbol{\psi}_{dummy}(t) \begin{bmatrix} \mathbb{1}(t)_{Mon-Thu} \\ \mathbb{1}(t)_{Fri} \\ \mathbb{1}(t)_{Sat} \\ \mathbb{1}(t)_{Sun} \\ \mathbb{1}(t)_{Hol} \\ \mathbb{1}(t)_{Bridge} \end{bmatrix} \tag{6.3}$$

The Fourier basis functions have an increasing frequency and they model three different seasonal patterns; yearly, weekly and daily. Thus, we have angular velocities on one or more of the following three forms

$$\omega_i = \frac{2\pi}{T_i}, \quad i \in [1,3] \tag{6.4}$$

where T_i are the different seasons (yearly, weekly, and daily) specified in hours (8760, 168, and 24 hours respectively).

The models were trained and evaluated using different numbers of polynomial bases, Fourier bases and external inputs.

6.1.1 Regularization

Elastic Net, Ridge and LASSO regularization were all conducted on the model set up defined in equation (6.1). This was done in order to prevent overfitting when introducing a high degree of model complexity as well as to identify relevant variables. Elastic Net was trained for different values of λ and α. Note that the penalty λ can also be written as a vector, giving different penalties to different parameters, where a λ value of zero means that there is no penalty. Finally, note that Ridge and LASSO are special cases of Elastic Net when $\alpha = 0$ and $\alpha = 1$.

6.2 Stochastic Time-Varying Model

The stochastic time-varying model was implemented to capture structural changes over time otherwise not found by the deterministic approach. A Kalman filter was used to recursively estimate the parameters over the training period, where the last estimation served as coefficients in the regression model to predict the ratios in the test set. The different sets of input variables retrieved from the best-performing deterministic models were chosen as input to this model.

Two variations of the Kalman filter and thus two stochastic model set ups were used. The first model set up was defined as follows

$$\begin{cases} y_t = \mathbf{w}_t^T \boldsymbol{\phi}(\mathbf{x}_t) + \sigma(\mathbf{x}_t) e_t \\ \mathbf{w}_t = \mathbf{w}_{t-1} + \boldsymbol{\epsilon}_t \end{cases} \tag{6.5}$$

where the parameters $\mathbf{w}_t$ represents the hidden state space that the Kalman filter estimates. Since we have observed that the month-to-hour ratios tend to have a higher variance during the winter than the summer, the variance function $\sigma(\mathbf{x}_t)$ was introduced

$$\sigma(\mathbf{x}_t) = 1 + A\cos\omega t \tag{6.6}$$

where $\omega = \frac{2\pi}{T_{year}}$ is the angular velocity per hour and A is the amplitude giving the seasonal difference of the measurement variance.

A detailed description of the filter is specified below. To start with, the prediction steps for the hidden state were

$$\hat{\mathbf{w}}_{t+1} = \hat{\mathbf{w}}_{t|t} \tag{6.7}$$

$$V[\hat{\mathbf{w}}_{t+1}] = \mathbf{P}_{t+1|t} = \mathbf{P}_{t|t} + \mathbf{Q} \tag{6.8}$$

and the update steps were as follows.

$$S_{t|t-1} = \boldsymbol{\phi}(\mathbf{x}_t)^T \mathbf{P}_{t|t-1} \boldsymbol{\phi}(\mathbf{x}_t) + R\sigma(\mathbf{x}_t) \tag{6.9}$$

$$\mathbf{K}_t = \mathbf{P}_{t|t-1} \boldsymbol{\phi}(\mathbf{x}_t) [S_{t|t-1}]^{-1} \tag{6.10}$$

$$\hat{\mathbf{w}}_{t|t} = \hat{\mathbf{w}}_{t|t-1} + \mathbf{K}_t (y_t - \boldsymbol{\phi}(\mathbf{x}_t)^T \hat{\mathbf{w}}_{t|t-1}) \tag{6.11}$$

$$\mathbf{P}_{t|t} = (\mathbf{I} - \mathbf{K}_t \boldsymbol{\phi}(\mathbf{x}_t)^T) \mathbf{P}_{t|t-1} \tag{6.12}$$

$$\tilde{y}_{t|t} = y_t - \boldsymbol{\phi}(\mathbf{x}_t)^T \hat{\mathbf{w}}_{t|t} \tag{6.13}$$

The initial parameter mean and variance was taken from the associated deterministic models.

When predicting the month-to-hour ratios with this model for the test period, it is not possible to update the parameters since we cannot compute the Kalman gain when we do not know what the observed values will be. Furthermore the prediction horizon for an HPFC can be up to five years and thus it might be problematic using these constant parameters.

With that in mind, we introduced a modified version of the stochastic model

$$\begin{cases} y_t = \mathbf{w}_t^T \boldsymbol{\phi}(\mathbf{x}_t) + \sigma(\mathbf{x}_t) e_t \\ \mathbf{w}_t - \overline{\mathbf{w}} = \mathbf{a}(\mathbf{w}_{t-1} - \overline{\mathbf{w}}) + \boldsymbol{\epsilon}_t \end{cases} \tag{6.14}$$

where the parameter vector $\mathbf{w}$ is assumed to follow an $AR(1)$-process, $\overline{\mathbf{w}}$ represents the mean vector of the Kalman parameter estimates and $\mathbf{a}$ is a vector with every element $|a_p| < 1$. This means that for each parameter a model defined as in equation (5.15) was estimated

$$w_t - \overline{w} = a(w_{t-1} - \overline{w}) + \epsilon_t \tag{6.15}$$

over the training period. These respective models were then used to predict each parameter over the test period. Thus, $\mathbf{w}$ varies as before over the training set, but for the test set the parameters are predicted with the autoregressive models. This will result in that the parameters over time will converge to their mean $\overline{\mathbf{w}}$.

6.3 Benchmark Model

Finally, a benchmark model was introduced as well. The model set up is the same as defined in equation (6.1). However, the benchmark does not contain any external inputs. It is only fitted against the Fourier bases and the dummy variables. It is introduced to measure the effects of including external regressors.

6.4 Model Configuration and Complexity

All models, except for the stochastic, were implemented in Python with the help of the *statsmodels* package, which provides functions for estimation of many statistical models and tools for statistical data exploration.

Before forming the basis functions and fitting the models, the fundamentals were standardized, i.e. transformed to being zero-mean and having a standard deviation of one. This was done partly to avoid multicollinearity for the polynomial bases and partly to avoid numerical issues in calculations.

The deterministic model was trained and tested for all combinations of external inputs, 7 yearly, 4 weekly and 10 daily Fourier bases. Furthermore, it was run for all polynomial order degrees up to 4. The reason for this specific configuration was due to seeing no improvements when introducing a higher degree of complexity in initial runs. Especially, it was found that the models tended to overfit for polynomials with a degree exceeding 4, and an addition of more Fourier bases did not generate better results. Finally, the benchmark model was chosen as the best-performing model without external inputs.

After some initial testing, the configuration for the regularization techniques were chosen as a trade-off between computational time and the ambition to examine more complex basis functions. The parameters varied were the penalty coefficient λ, 40 steps between -5 and 1 on a logarithmic scale and the $l1$-ratio where $\alpha \in \{0, 0.05, ..., 0.95, 1.00\}$. For combinations with good performance an additional run was made varying the parameters over a smaller surrounding interval. For the Elastic Net, the polynomial bases were extended to the 10th degree for all fundamental variables. Also the Fourier bases were set to the 10th multiple of the respective angular velocity. Restrictions on the basis degrees were considered necessary due to computational complexity.

For the Kalman filter the variances of the measurement noise R and state noise $\mathbf{Q}$ had to be chosen. The model was trained with both variances ranging from -5 to 0 on a logarithmic scale. In addition the amplitude A took the values of $A \in [0, 0.2, 0.4, 0.6, 0.8]$.

6.5 Model Evaluation and Selection

When evaluating our models and selecting the one with the best performance, we have resorted to using both validation metrics and the criteria presented in Section 3.3. The validation metrics used are the following: RMSE, MAE, R^2 and the price prediction errors introduced in Section 5.5.4. The best-performing models are firstly narrowed down using the validation metrics and the best one among these will be selected after validating against the criteria in Section 3.3 by graphically inspecting the out-of-sample predictions.

Chapter 7

Results

7.1 Deterministic Time-Varying Model

The two tables below summarizes the best results. Table 7.1 specifies the model configuration, i.e. external inputs, polynomial degree order and the number of Fourier bases with yearly, weekly and daily frequencies. Table 7.2 gives the out-of-sample validation metrics for the corresponding models. For each validation metric, the model with the best performance is presented.

Model	**Externals**	**Polynomial Order**	**Yearly Freq.**	**Weekly Freq.**	**Daily Freq.**
A	Cross-Border Flow Installed Wind Capacity Consumption	1	5	0	1
B	Cross-Border Flow Installed Wind Capacity Consumption	1	1	0	1
C	Cross-Border Flow Installed Wind Capacity Consumption Temperature Deviation Hydrological Balance	1	5	1	10
Benchmark	N/A	1	5	4	10

Table 7.1: Model specifications for the best-performing deterministic models and the benchmark model. Information regarding the external inputs used, polynomial order degree, and the number of different Fourier bases with yearly, weekly, and daily frequencies, for each model is summarized.

Model	RMSE	MAE	R^2	Absolute Error	Squared Error
A	**0.11968**	0.079445	**0.42673**	2.9810	20.665
B	0.12021	**0.079299**	0.42171	**2.9674**	20.791
C	0.11979	0.080237	0.42570	3.0009	**20.584**
Benchmark	0.13314	0.092093	0.29062	3.4264	25.039

Table 7.2: Out-of-sample validation metrics for the best-performing deterministic models and the benchmark model. For each validation metric, the best one has been highlighted in bold font.

Now finally, validating against the criteria given in Section 3.3, Model C was deemed as the best one. It

illustrates the desired seasonalities as well as being robust against outliers and overfitting. These characteristics are displayed in Figures 7.1, 7.2 and 7.3.

Looking at Figure 7.1, it shows that the model captures the overall movement while being robust against the extreme ratios characteristic for the winter months. However, the model seems to be lacking in that it displays nearly a constant amplitude despite time of the year, with the true ratios displaying more variation. Attempts were made for Model C to include an additional basis expansion, multiplying the fundamentals with yearly sinusoidals, to compensate for this. However, the results were unsatisfactory. Moreover, the model tends to predict too low ratios in the weekend nights whereas mid-day ratios Monday to Friday in the warmer months are set at a higher level than desired, as seen in Figure 7.3.

The intra-yearly periodicities are even more clearly illustrated in Figure 7.3 with peakier ratios during the winter than during the summer. However, the model does not perform optimally in regards to modelling the difference between off-peak and peak hours during summer months as seen in the figure, which was a challenge throughout this study.

Finally, the intra-weekly and intra-daily seasonalities are displayed in Figure 7.2. The model performs well in these regards. It differentiates between weekdays and weekend days and predicts the pronounced morning and evening peaks well. However, the model has difficulties in modelling the low-demand hours on non-working days. Adding new dummy variables for these morning periods was tested for Model C but with unsatisfying results.

To compare the impact of the different parameters, the most influential coefficients of the model with normalized inputs have been illustrated in Figure A.4 in Appendix A. Among the fundamentals, consumption has by far the strongest impact on the modelled ratios. Since the consumption follows the same seasonal pattern as the spot prices and very much reflects the demand in the market, this is not surprising. Temperature deviation and cross-border flow are other strong factors affecting the modelled ratios, but not nearly to the extent of consumption. Moreover, large parameter values can be observed connected to several Fourier basis functions, indicating strong seasonality, and categorical variables for Sundays and Holidays. Here it should be noted that Monday-Thursday serves as the reference variable. A final observation is that these models greatly outperforms the benchmark model according to both metrics and the seasonality patterns illustrated by the models. The validation metrics for the benchmark model is presented in Table 7.2 and its seasonality patterns are illustrated in Figures A.1, A.2 and A.3 in Appendix A. This suggests that an inclusion of external inputs certainly improves the model in this particular model set up.

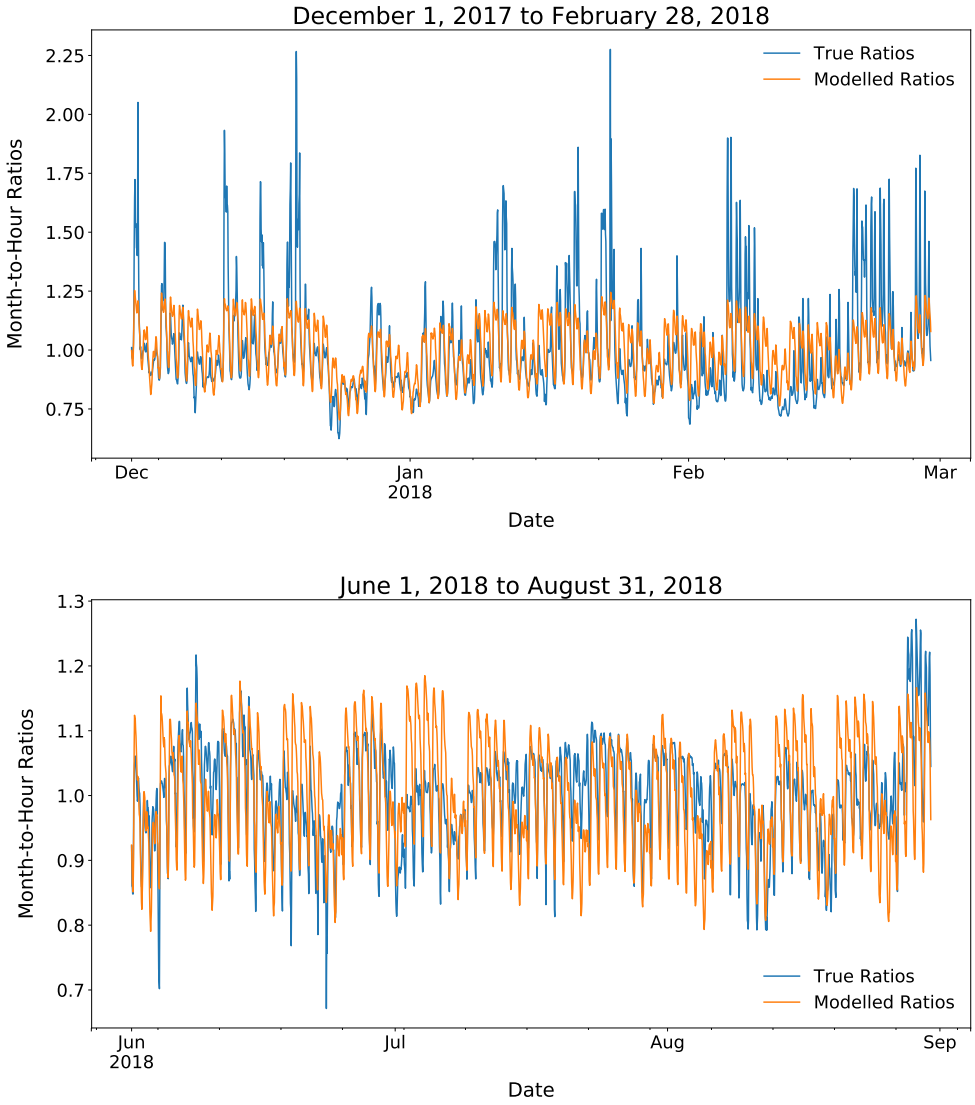


Figure 7.1: True and modelled ratios over winter/summer months for the best-performing deterministic model.

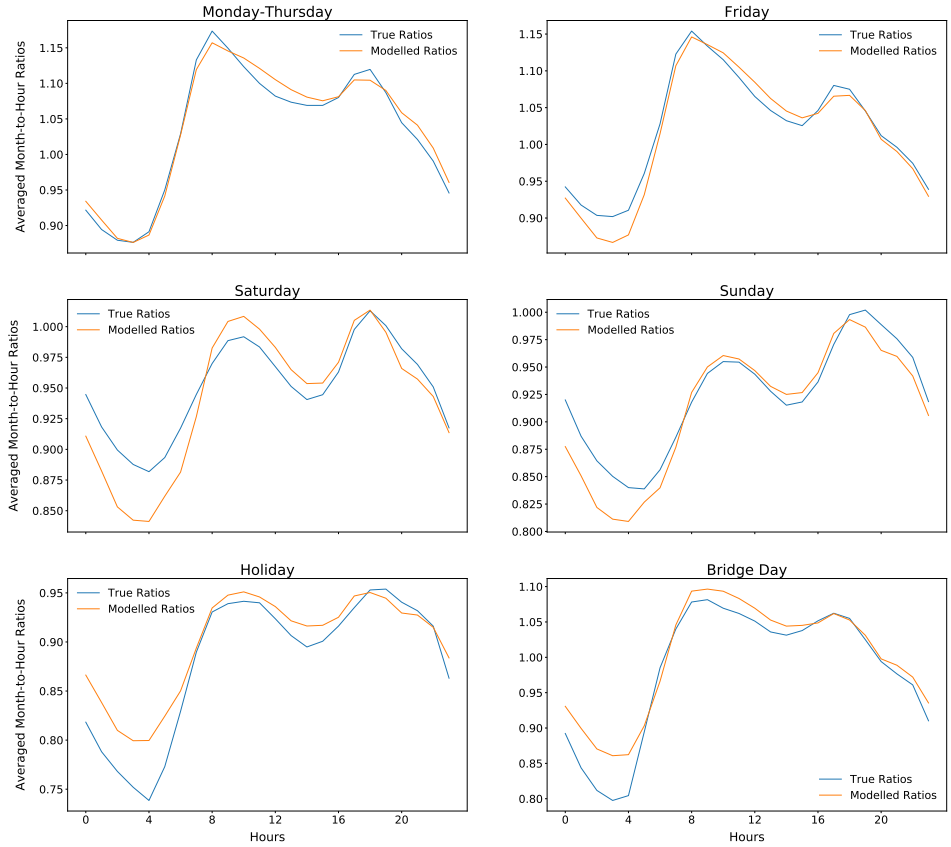
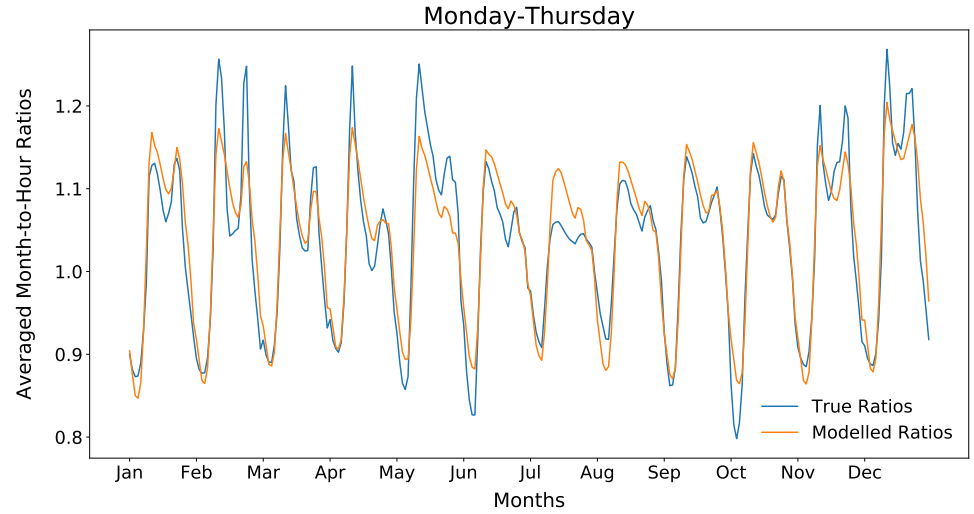
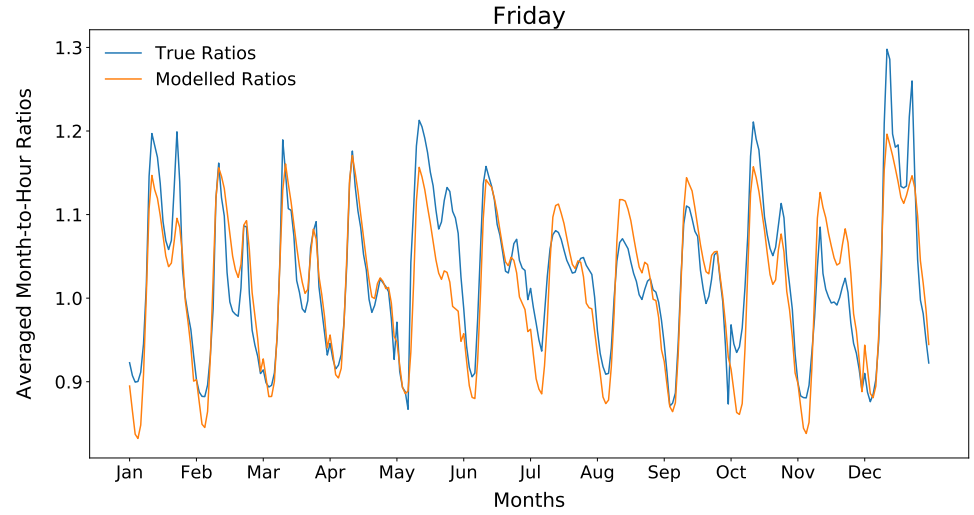
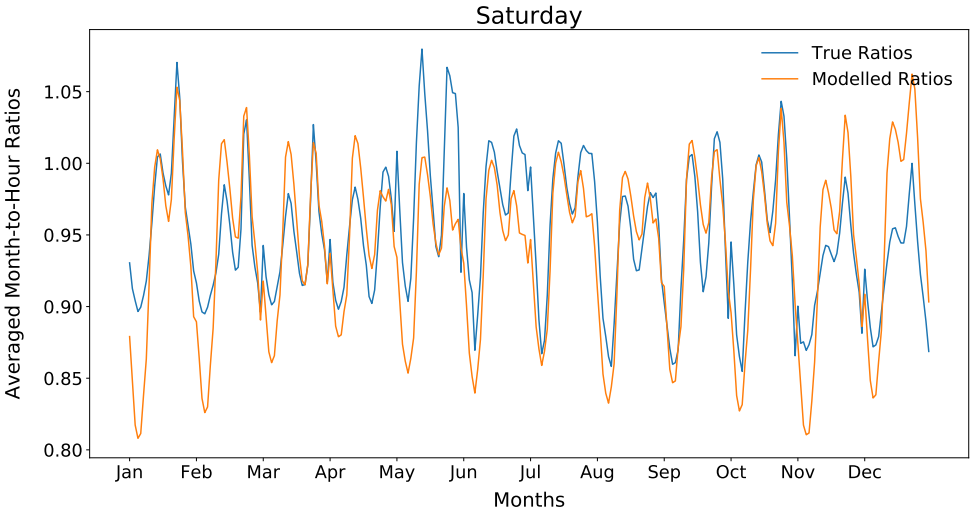


Figure 7.2: Intra-daily seasonalities per day type for the best-performing deterministic model. Month-to-hour ratios averaged for each day type for every hour of the day over the test period.


Monday-Thursday
Averaged Month-to-Hour Ratios
1.2
1.1
1.0
0.9
0.8
True Ratios
Modelled Ratios
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Months


Friday
Averaged Month-to-Hour Ratios
1.3
1.2
1.1
1.0
0.9
True Ratios
Modelled Ratios
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Months


Saturday
Averaged Month-to-Hour Ratios
1.05
1.00
0.95
0.90
0.85
0.80
True Ratios
Modelled Ratios
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Months

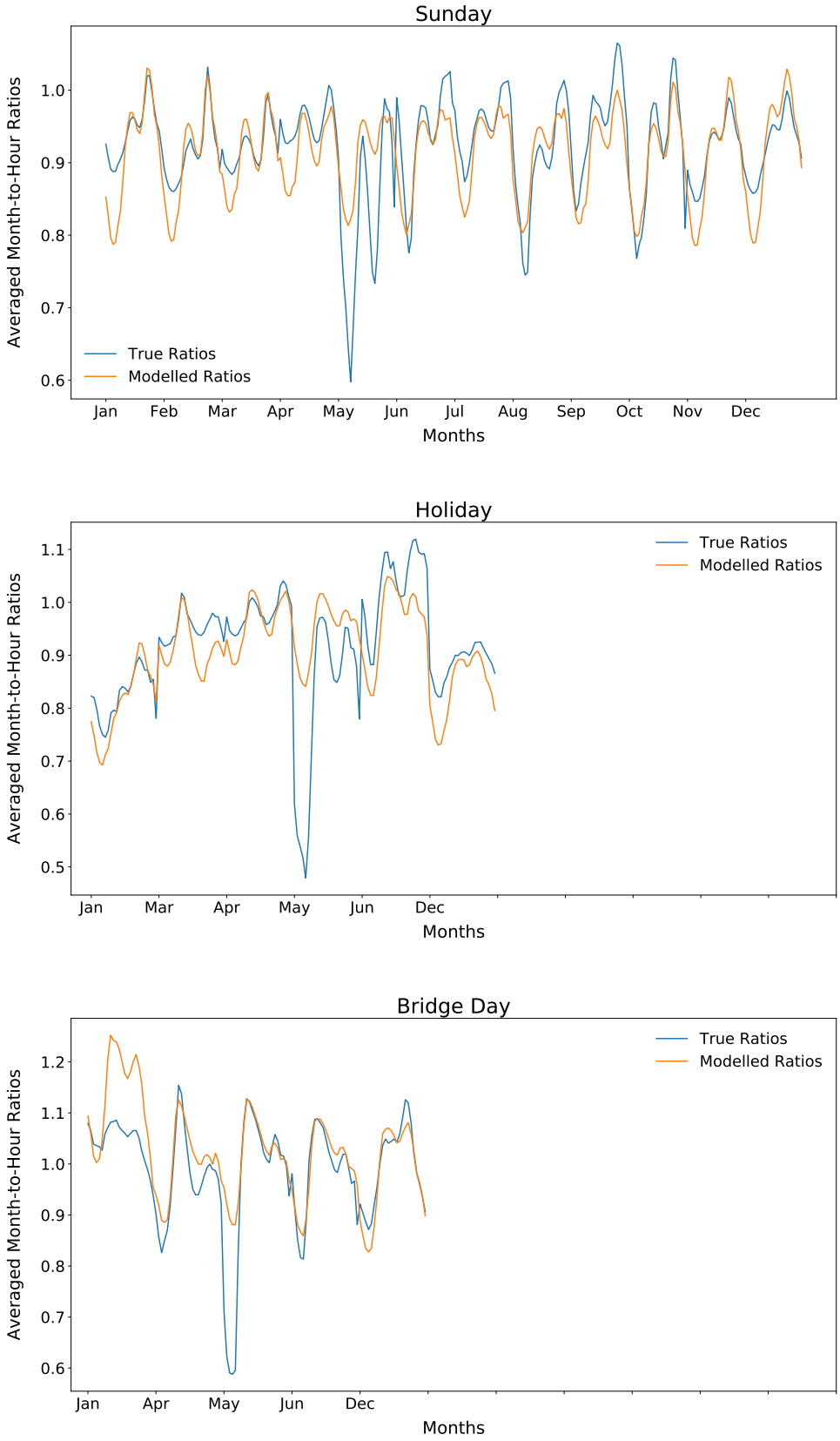


Figure 7.3: Intra-yearly seasonalities for the best-performing deterministic model. Month-to-hour ratios averaged for each day type and month for every hour of the day over the test period, starting with the hours of January and ending with the hours of December.

7.2 Regularization

This section presents the results from the regularized regression techniques. Choosing the best model for each validation metric rendered in four different configurations presented in Table 7.3.

Model	λ	α	RMSE	MAE	R^2	Absolute Error	Squared Error
D	0.0030	0.80	**0.12139**	0.080636	**0.41025**	3.0332	21.276
E	0.0045	1.00	0.12414	**0.080081**	0.38324	3.0022	22.224
F	0.0050	0.95	0.12465	0.080099	0.37815	**3.0009**	22.409
G	0.0025	1.00	0.12140	0.080565	0.41020	3.0304	**21.274**
Benchmark	N/A	N/A	0.13314	0.092093	0.29062	3.4264	25.039

Table 7.3: Out-of-sample validation metrics for the best-performing regularized models and the benchmark model. For each validation metric, the best one has been highlighted in bold font.

After validating against the additional criteria, Model D was deemed as the best out of the four aforementioned models. It illustrates visible seasonal patterns and is robust against extreme ratios and thereby fulfills, to some extent, the quality measurements for a shape vector. However, it does severely underperform compared to the deterministic models when further studying its seasonalities in Figures 7.4, 7.5 and 7.6.

First of all, when studying the resulting intra-daily seasonalities for the regularized model in Figure 7.5, it displays worse results compared to the deterministic approaches. The model displays a more rounded contour and it underperforms in regards to capturing the peak hours of the different day types. It especially underperforms for the day type Holiday. However, when analyzing the coefficients of the regularized models, it was observed that all the day type coefficients had been set to zero. This might explain the poor intra-daily seasonality displayed by Holiday. In order to improve the intra-daily performance, experiments were conducted where the penalty for different variables were removed. Modelling attempts were made where the penalty for the daily Fourier bases and the day type dummy variables were set to zero. In one attempt, the penalty term for the daily Fourier bases were set to zero, in the second attempt the penalty for the dummy variables were set to zero, and in a third attempt the penalties for both the dummy variables and the daily Fourier bases were removed simultaneously. This was done in three separate runs for all variations of our prespecified α and λ values. However, no improvements were observed in these attempts.

Moreover, studying Figure 7.4, similar results as the deterministic models are retrieved for the illustrated winter months. The regularized model however displays more of an unstable behavior during the selected summer period with a sudden variation, which is not quite desirable and was not illustrated by the previous deterministic model.

Furthermore in Figure 7.6, the problems regarding the overall amplitude of the curve continues. Studying the averaged ratios for Monday-Thursday, the amplitude seems to be lower late-year than during the summer, although the true ratios show the opposite behaviour.

Since Elastic Net adds a penalty term on the parameters, it is of interest to further investigate the estimated coefficients to see which were deemed as relevant. These are illustrated in Figure A.5. The highest parameters are those connected to the yearly sinusoidals. The most significant fundamentals are cross-border flow of the first power and consumption of the third power.

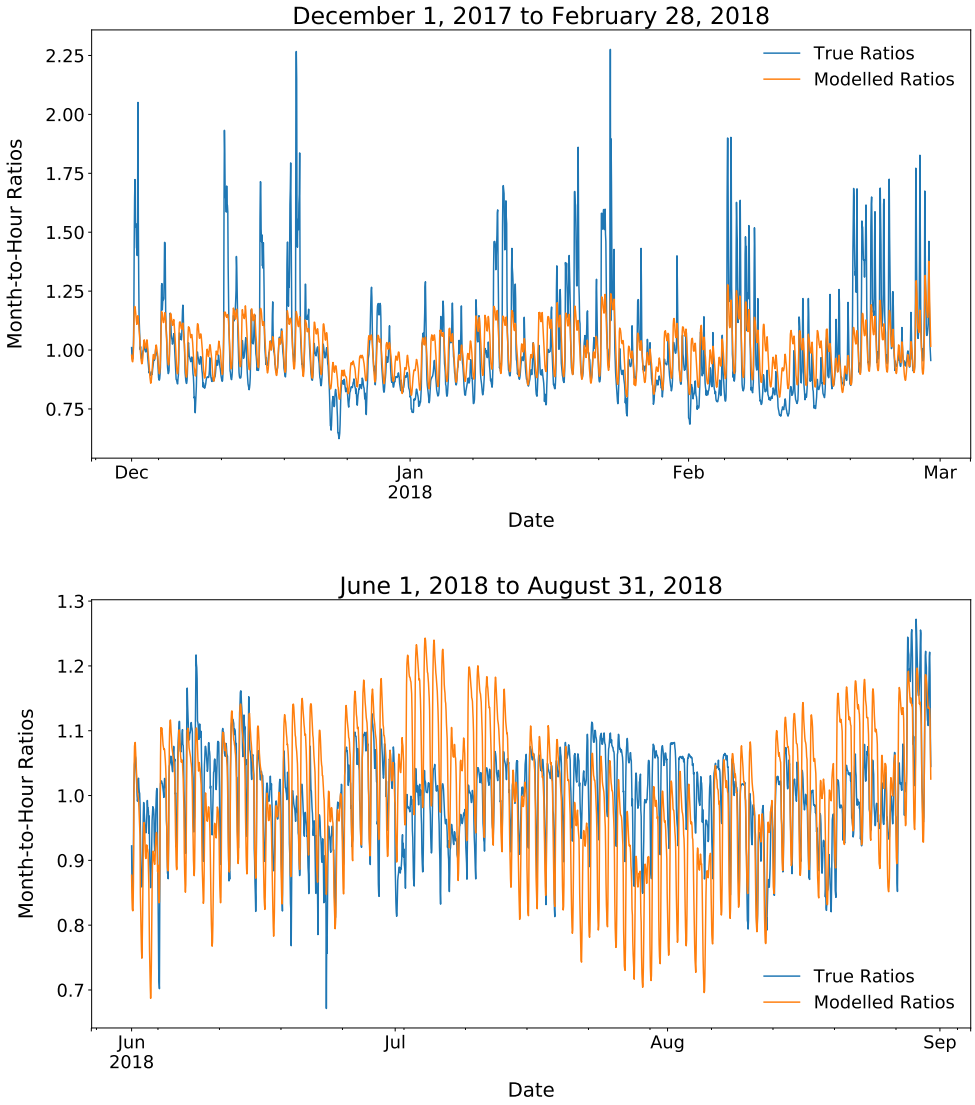


Figure 7.4: True and modelled ratios over winter/summer months for the best-performing regularized model.

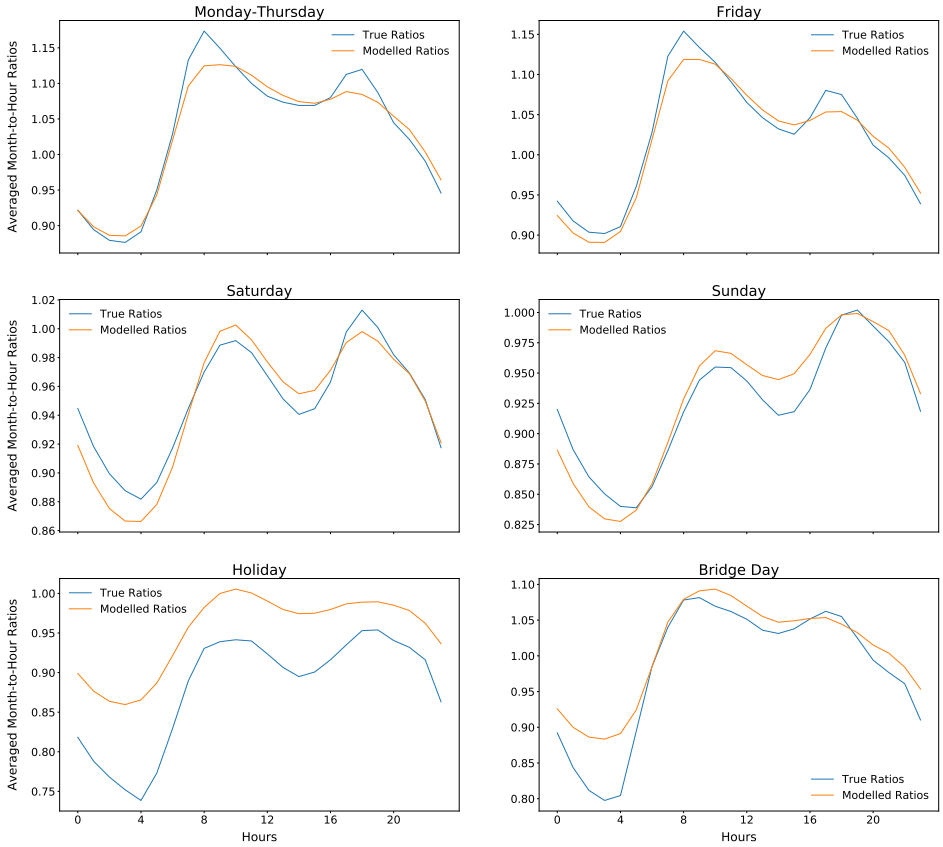
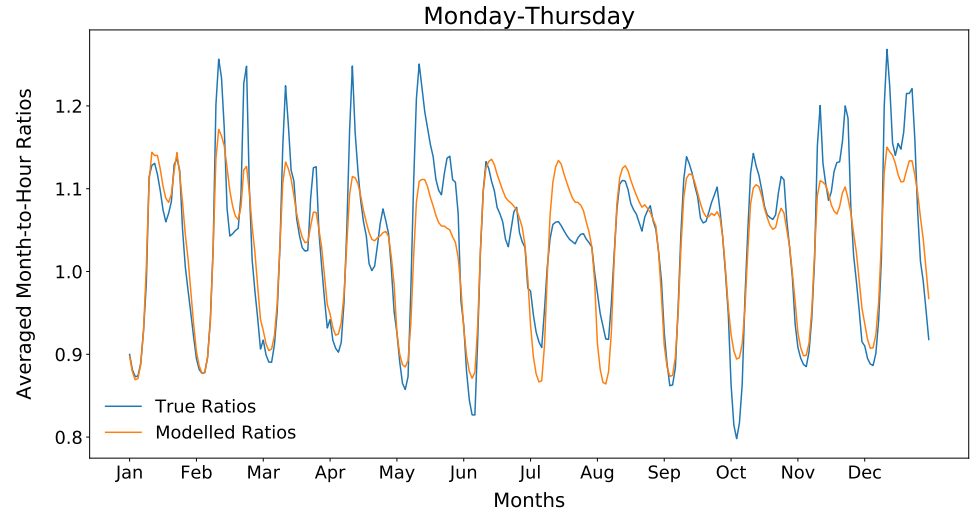
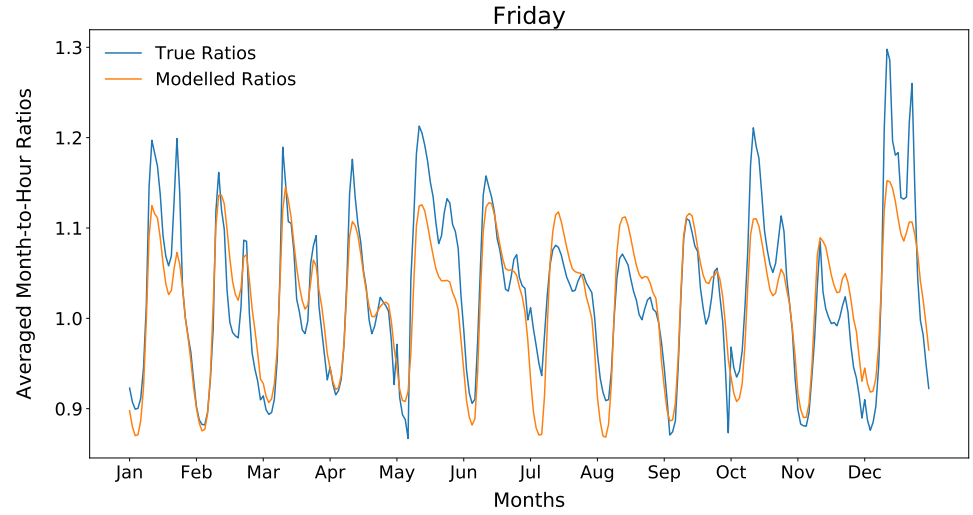
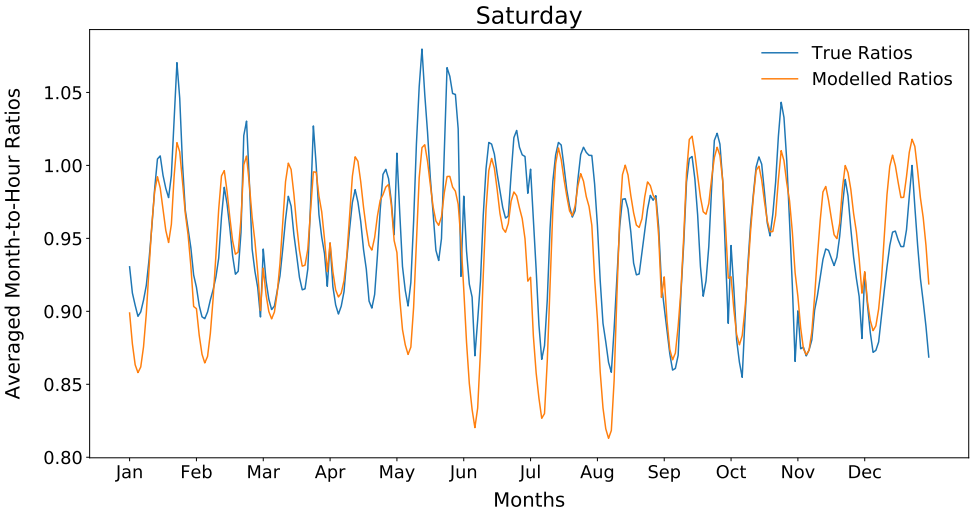


Figure 7.5: Intra-daily seasonalities per day type for the best-performing regularized model. Month-to-hour ratios averaged for each day type for every hour of the day over the test period.


Monday-Thursday
Averaged Month-to-Hour Ratios
1.2
1.1
1.0
0.9
0.8
True Ratios
Modelled Ratios
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Months


Friday
Averaged Month-to-Hour Ratios
1.3
1.2
1.1
1.0
0.9
True Ratios
Modelled Ratios
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Months


Saturday
Averaged Month-to-Hour Ratios
1.05
1.00
0.95
0.90
0.85
0.80
True Ratios
Modelled Ratios
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Months

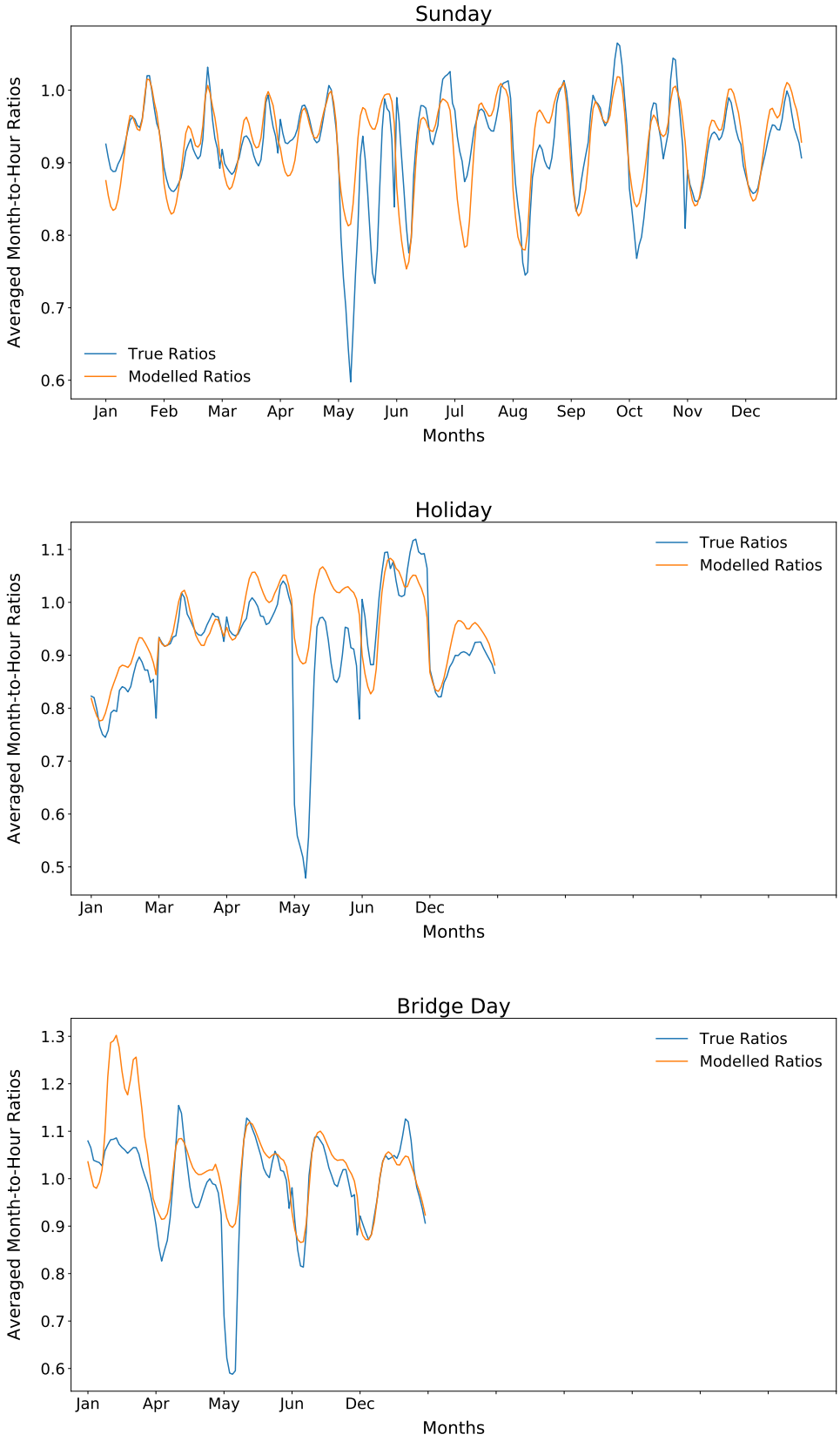


Figure 7.6: Intra-yearly seasonalities for the best-performing regularized model. Month-to-hour ratios averaged for each day type and month for every hour of the day over the test period, starting with the hours of January and ending with the hours of December.

7.3 Stochastic Time-Varying Model

The stochastic time-varying models were implemented using the two different set ups defined in equations (6.5) and (6.14). The first model assumes flat parameters and the second one assumes that the parameters follow an AR(1). Models with *Flat* and *AR(1)* parameters respectively were trained using the three different model configurations (A, B and C) presented in Table 7.1, resulting in a total of 6 different model configurations trained with varying filter specifications.

The model specifications for the best models are presented in Table 7.4 with their respective Kalman filter specifications summarized in Table 7.5. Table 7.6 gives the out-of-sample validation metrics for the corresponding models. For each validation metric, the model with the best performance is presented.

Model	**Parameter**	**Externals**	**Polynomial Order**	**Yearly Freq.**	**Weekly Freq.**	**Daily Freq.**
H	Flat	Cross-Border Flow Installed Wind Capacity Consumption Temperature Hydrological Balance	1	5	1	10
I	AR(1)	Cross-Border Flow Installed Wind Capacity Consumption	1	1	0	1
Benchmark	N/A	N/A	1	5	4	10

Table 7.4: Model specifications for the best-performing stochastic models and the benchmark model.

Model	**Measurement Noise Variance**	**State Noise Variance**	**Variance Function Amplitude**
H	0.00599	0.00001	0.8
I	0.00001	0.00046	0.0

Table 7.5: Kalman filter specifications for the best-performing stochastic models.

Model	RMSE	MAE	R^2	Absolute Error	Squared Error
H	**0.11798**	0.080142	**0.44289**	2.9983	**19.889**
I	0.12113	**0.076264**	0.41273	**2.8590**	21.117
Benchmark	0.13314	0.092093	0.29062	3.4264	25.039

Table 7.6: Out-of-sample validation metrics for the best-performing stochastic models and the benchmark model. For each validation metric, the best one has been highlighted in bold font.

Now validating against the additional criteria, Model H was established as the best one out of the two presented. It meets the criteria of an adequate shape vector. However, it does not present any major improvements compared to our previous deterministic models, even though it has slighty better performance metrics. Nevertheless, it outperforms the regularized techniques in all regards.

The goal of applying the Kalman filter is that the model will be able to adapt to changes in how the regressors impact the ratios over time. Judging from the Model H metrics, in regards to RMSE, R^2 and Squared Error, it seems as if its predictive performance has improved compared to the deterministic and regularized approaches. It also outperformed the otherwise similar model with parameters that follow an AR(1)-process. Since AR(1) parameters will return to their mean, this outperformance might suggest that the structural changes in the relationship between the variables are valid for the test period as well, making the Flat parameters more appropriate.

In Figure 7.7, 7.8 and 7.9 on the following pages, Model H is plotted against the true ratios. Starting with the intra-yearly cycle and studying the two sample plots in Figure 7.7, there is not much of a difference compared to our best model up to this point, the deterministic linear regression model introduced previously, i.e. Model C. Furthermore, in this figure it is also observed that the model is robust to extreme ratios and follows the overall trend of the true ratios.

Studying the intra-daily performance in Figure 7.8, we conclude that it captures the overall pattern well. The stochastic model seems to predict more distinct morning and evening peaks on Monday-Thursdays and Saturdays. However, the issue regarding inaccurate and large spreads for the night and morning periods still remain. Since only these small differences exists, it is difficult to draw any general conclusions from this comparison.

Finally, when inspecting the averaged month-to-hour ratios over the year in Figure 7.9, we observe the same general pattern as for the earlier deterministic models. The amplitude of the shape is too large on weekdays during the summer months and on weekend days during the winter months. We also observe slightly more distinct peaks for the working days here than for our deterministic model.

As for the coefficients, illustrated in Figure A.6, the yearly cosine function with lowest frequency ($\cos(1\omega_{year}t)$) and consumption has the largest impact on the predictions. An interesting observation is that the parameters of the daily sinusoidals with lowest frequency ($\cos(1\omega_{day}t)$, $\sin(1\omega_{day}t)$) have a larger impact here than for the other models, suggesting that the daily seasonality has become more significant in recent years.

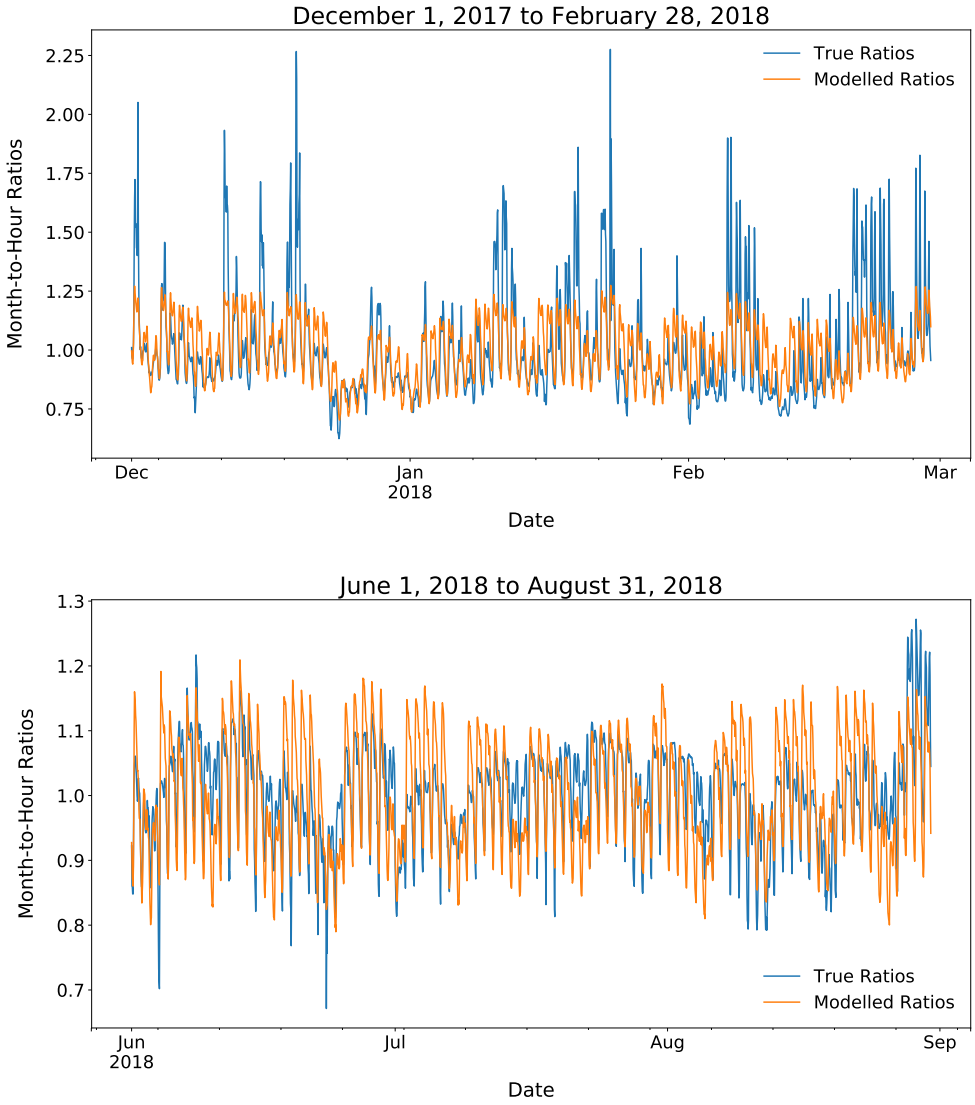


Figure 7.7: True and modelled ratios over winter/summer months for the best-performing stochastic model.

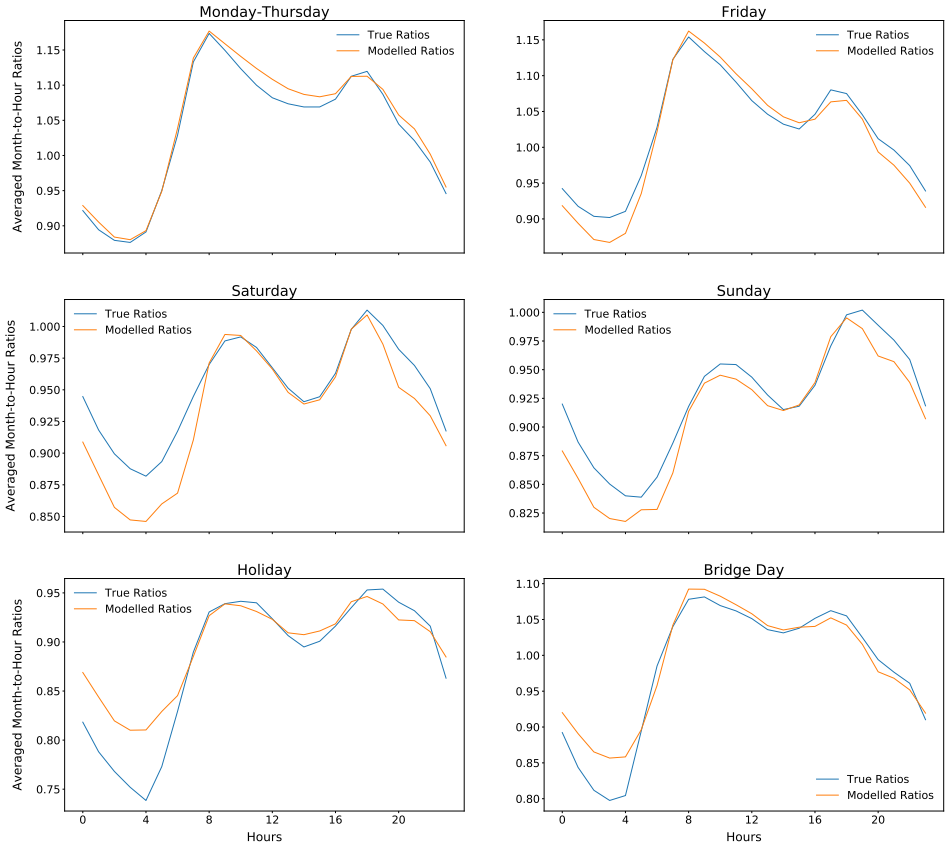
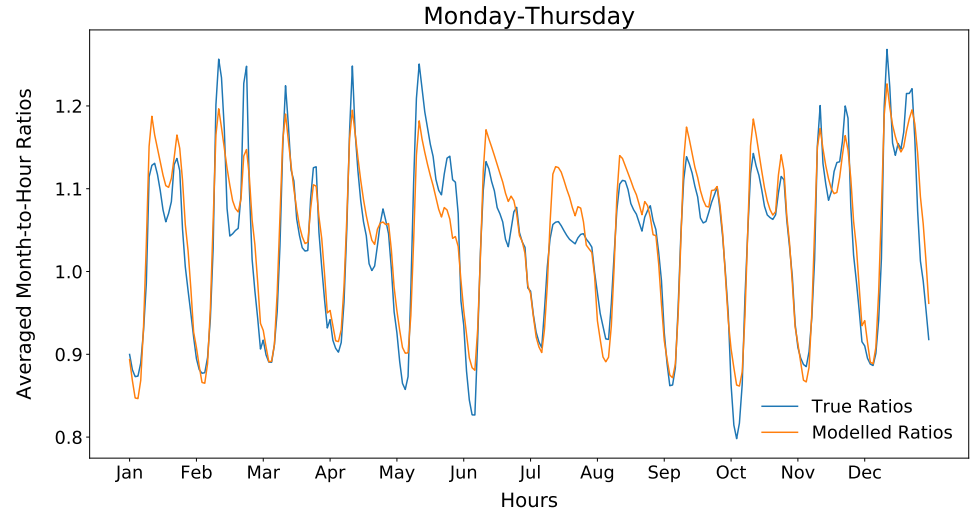
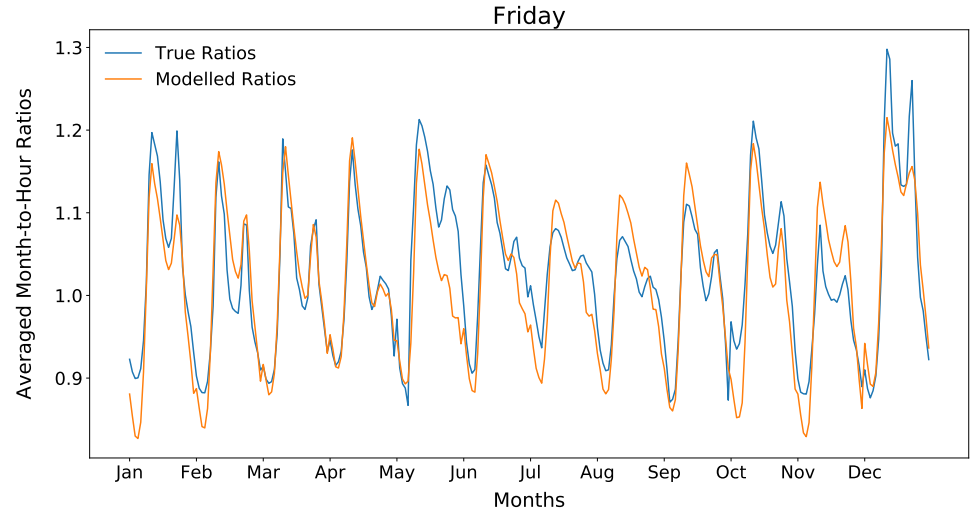
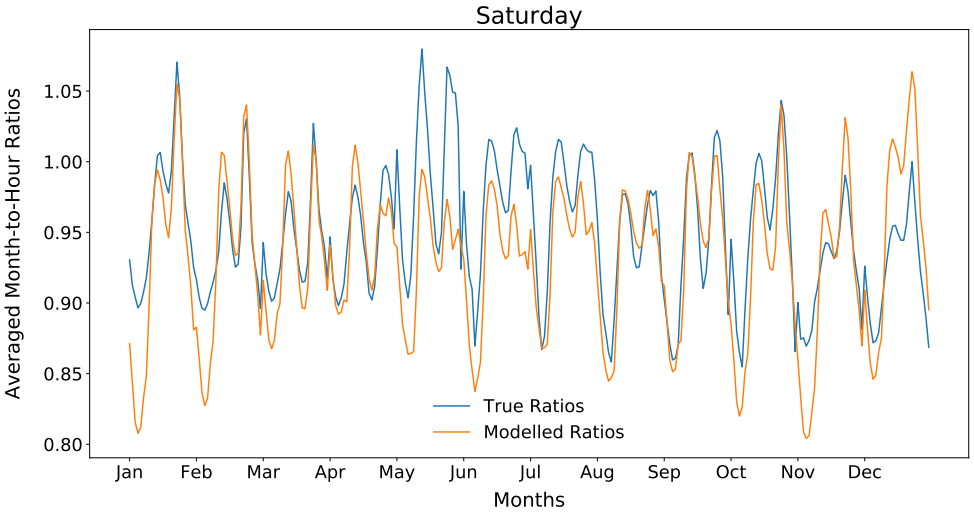


Figure 7.8: Intra-daily seasonalities per day type for the best-performing stochastic model. Month-to-hour ratios averaged for each day-type for every hour of the day over the test period.


Monday-Thursday
Averaged Month-to-Hour Ratios
1.2
1.1
1.0
0.9
0.8
True Ratios
Modelled Ratios
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Hours


Friday
Averaged Month-to-Hour Ratios
1.3
1.2
1.1
1.0
0.9
True Ratios
Modelled Ratios
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Months


Saturday
Averaged Month-to-Hour Ratios
1.05
1.00
0.95
0.90
0.85
0.80
True Ratios
Modelled Ratios
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Months

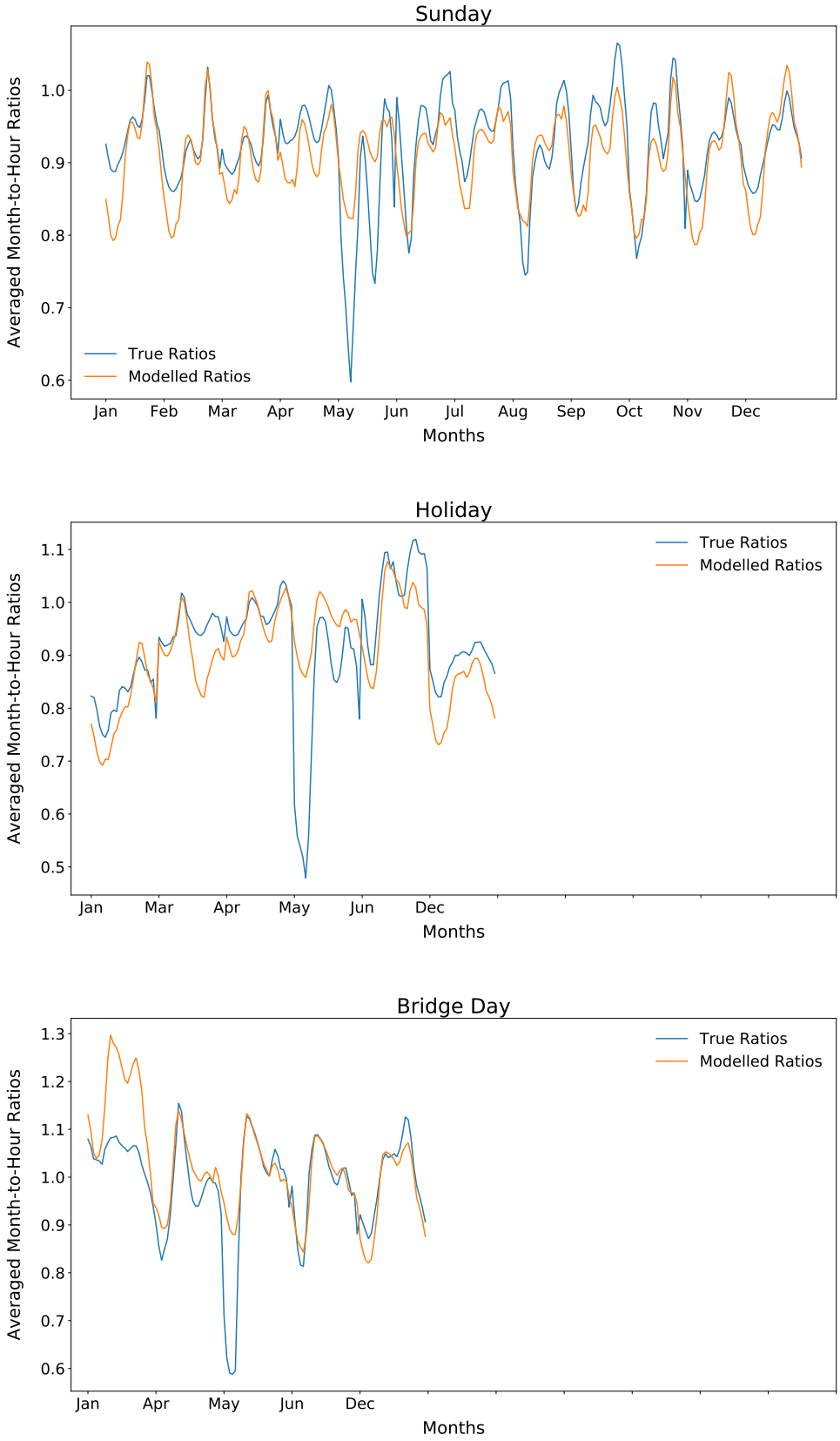


Figure 7.9: Intra-yearly seasonalities for the best-performing stochastic model. Month-to-hour ratios averaged for each day-type and month for every hour of the day over the test period, starting with the hours of January and ending with the hours of December.

Chapter 8

Conclusions

The purpose of this thesis was to investigate whether or not it is possible to employ a parametric approach, using external inputs, to estimate the month-to-hour ratios for the Nord Pool system price. This paper uses regression models with polynomial and Fourier basis functions, combined with regularization techniques and time series models. The following external inputs were examined: consumption, hydrological balance, temperature deviation, cross-border flow and installed wind capacity. When comparing the constructed models to a benchmark, which does not include any external inputs, it is concluded that parametric models with external regressors are well-suited for this estimation problem. Moreover, consumption was found to be one of the most influential explanatory variables among all of our top-performing models.

The deterministic and stochastic regression models gave similar results, even though, the metrics for the stochastic model was slightly better. However, the results are nearly identical when visualizing the seasonalities in the plots. Both of the models perform well and capture the seasonalities of the electricity spot price while being robust against extremities, which was the aim of the study. Furthermore, it differentiates between the yearly seasons and follows the distinguished off-peak/peak hour pattern. Our model does well in recognizing the day types configured through dummy variables, but there are deviations regarding the morning levels for Saturdays, Sundays, Holidays and Bridge Days. Attempts were made to model these periods with additional dummy variables, but the results were unsuccessful. Regarding the regularization techniques, they did not generate models outperforming the aforementioned. The intra-daily seasonalities in particular were not captured as well as in the deterministic and stochastic models. Another challenging aspect regarding the modelling, was to capture the summer trend. Our models generally illustrated more spikes during this period.

An additional comment on the stochastic time-varying model is that a rather simple function was used when varying the measurement noise over the year. Nevertheless, it gave satisfying results increasing the performance of the model. Using more refined techniques could potentially make it even better.

In this study, we have predicted the month-to-hour ratios using the actual outcomes of the different fundamentals. However, in a real-world application, one does not have access to this data and needs to resort to projected values instead. This will surely give different results. However, in this study we have used data that is extensively used by energy companies and that is possible to project into the future.

Furthermore, an interesting observation is that our top-performing models included installed wind capacity as their external regressor, despite it being non-stationary. Generally, care should be taken when including non-stationary data in a regression analysis. However, as mentioned earlier, this variable was of interest for us and we did not observe any negative impacts of including it.

Moreover, the topic of validation has been challenging throughout this research. As previously mentioned, our goal is not to construct a future price forecast and it is therefore not optimal to validate the shape vector against spot prices. However, there is yet no straightforward validation framework regarding the HPFC construction. Thus, we needed to resort to validation metrics comparing against realized

spot prices and to graphically validate our curve against important criteria inherent for the HPFC. In our case, the so-called graphical validation was quite important since the top-performing models were quite similar in the error measures. This might be a result of our target variable varying in a narrow interval for the entire data set, $y_t \in [0, 2.8]$, and thereby our metrics may lose some validity. Thus, our validation framework is not optimal and needs to be refined, but due to lack of other options, it worked well in our case. Moreover, we also found that the Mean Absolute Error (MAE) was not able to identify the best models. Among the models with the lowest MAE, some of them showed undesirable graphical traits, not illustrated by models with higher MAE.

Chapter 9

Further Work

For future research there are several areas that would be of interest to further investigate. To begin with, this thesis solely considers the month-to-hour ratios. Naturally, modifying this model to predict the complete shape vector with year-to-hour ratios would be of great interest. Moreover, to conduct a price calibration on the modelled seasonal shape in order to construct the complete HPFC would make for a more valid valuation since the HPFC is the final product of the shape vector.

Furthermore, more research regarding refined validation metrics and framworks are needed. When reviewing literature on HPFC modelling the authors mostly evaluate the modelled curves from some special visual characteristics. This valuation can be quite ambiguous in some cases. It is intuitive to see if a curve follows desired criteria or not, but when comparing several models with each other it can be difficult to distinguish differences in performance. Therefore, a more thorough review of this topic would be of great value for improving the investigative work of this thesis.

Another aspect of the thesis is the external regressors that were included. The portion of power from renewable energy sources are increasing rapidly and therefore it is interesting to study its effect on the HPFC. Although wind power capacity was included in the study, the wind power production would most likely provide the model with more relevant information. However, due to scarce data availability, wind power production needed to be set aside. Furthermore solar power will have an impact in the future as well, but up until now it is a small part of the Nord Pool supply and for that reason it was discarded in this study. The German market does not have any of these data limitations and for that reason we believe it would be of worthwhile to develop a modified version of our models on that market.

Appendix A

Appendix

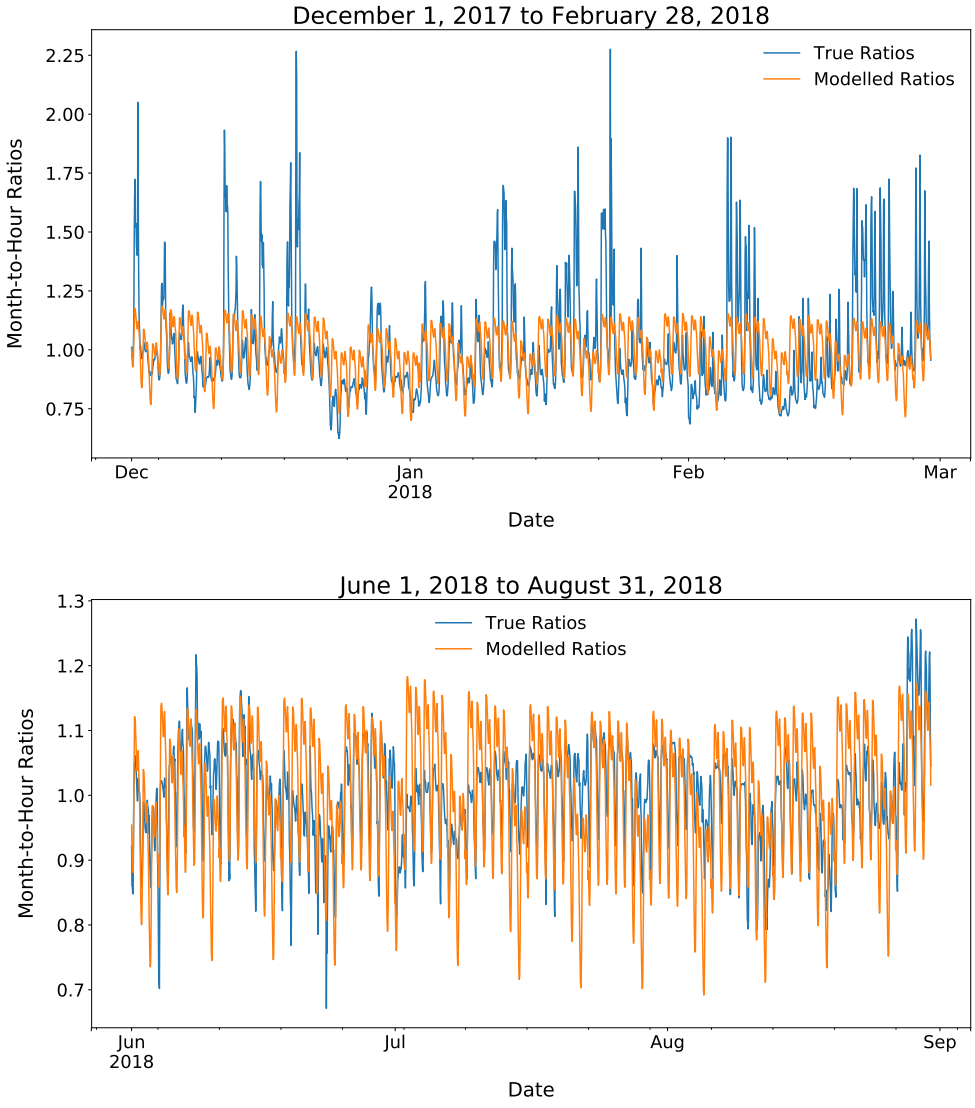


Figure A.1: True and modelled ratios over winter/summer months for the benchmark model.

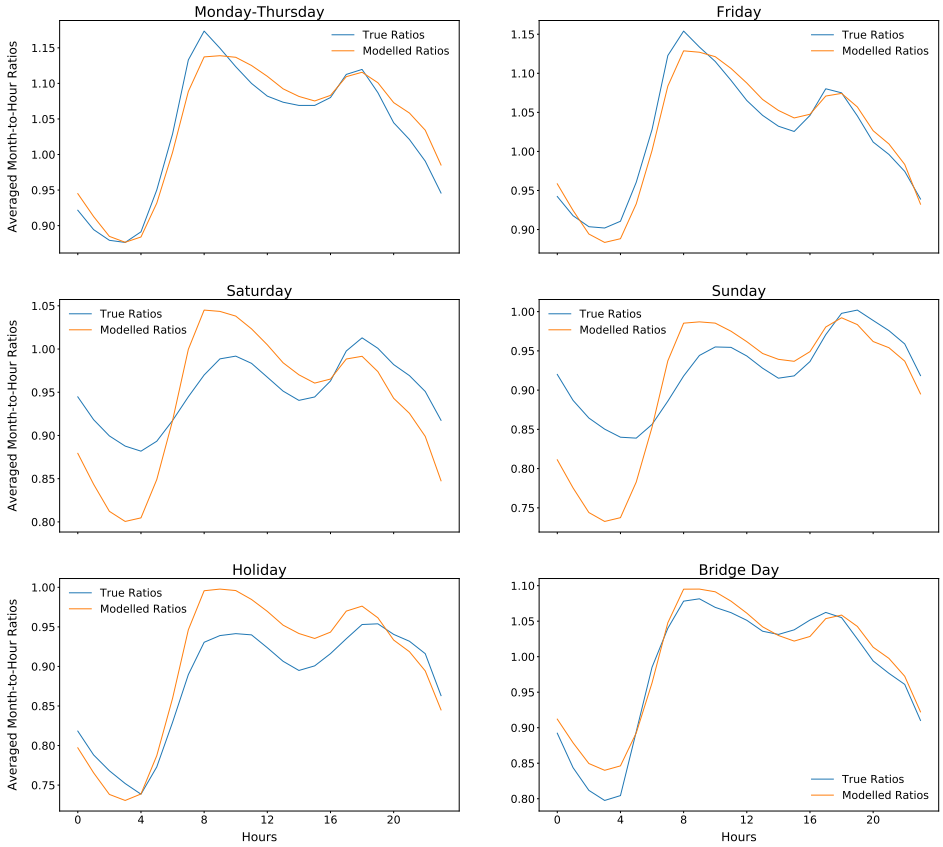
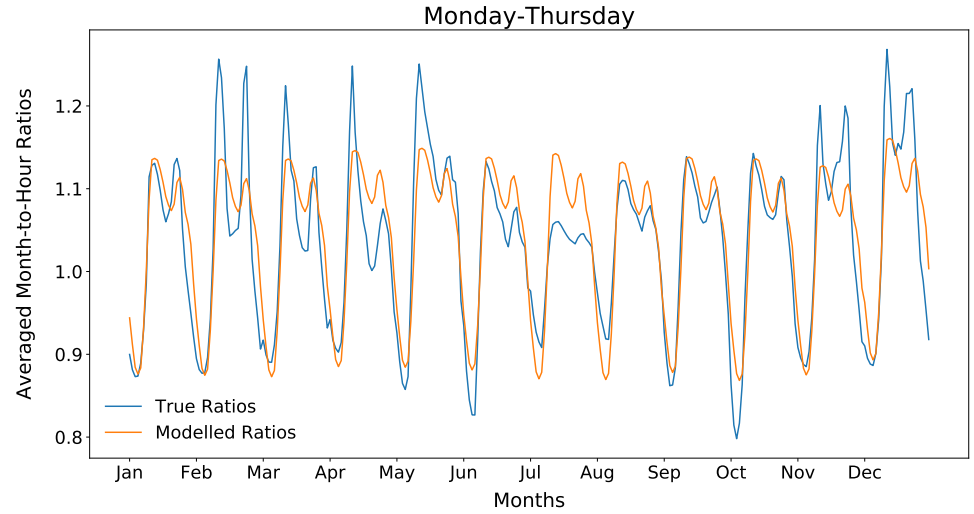
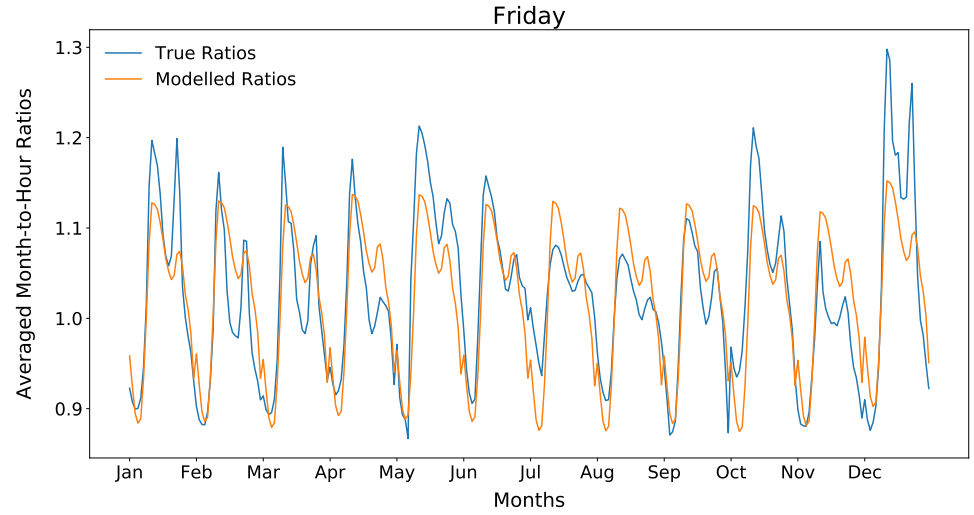
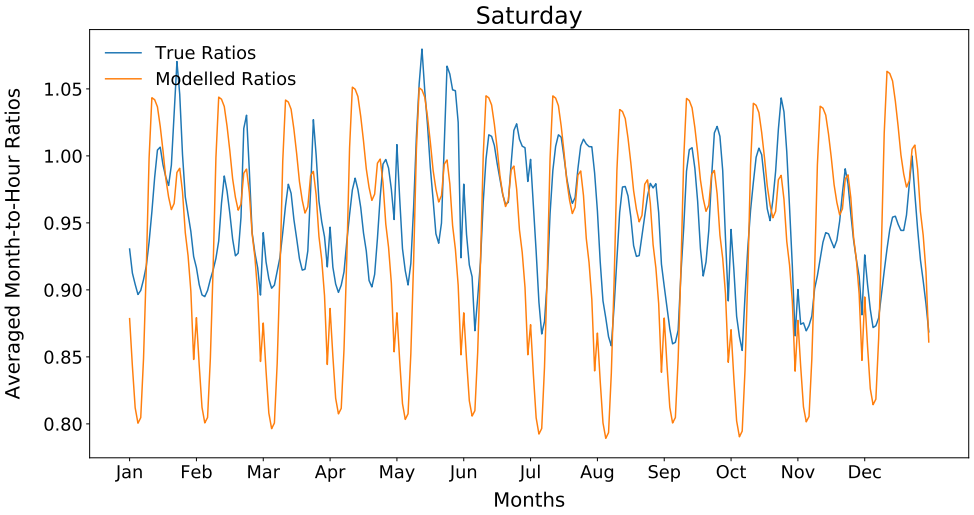


Figure A.2: Intra-daily seasonalities per day type for the benchmark model. Month-to-hour ratios averaged for each day-type for every hour of the day over the test period.


Monday-Thursday
Averaged Month-to-Hour Ratios
1.2
1.1
1.0
0.9
0.8
True Ratios
Modelled Ratios
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Months


Friday
Averaged Month-to-Hour Ratios
1.3
1.2
1.1
1.0
0.9
True Ratios
Modelled Ratios
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Months


Saturday
Averaged Month-to-Hour Ratios
1.05
1.00
0.95
0.90
0.85
0.80
True Ratios
Modelled Ratios
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Months

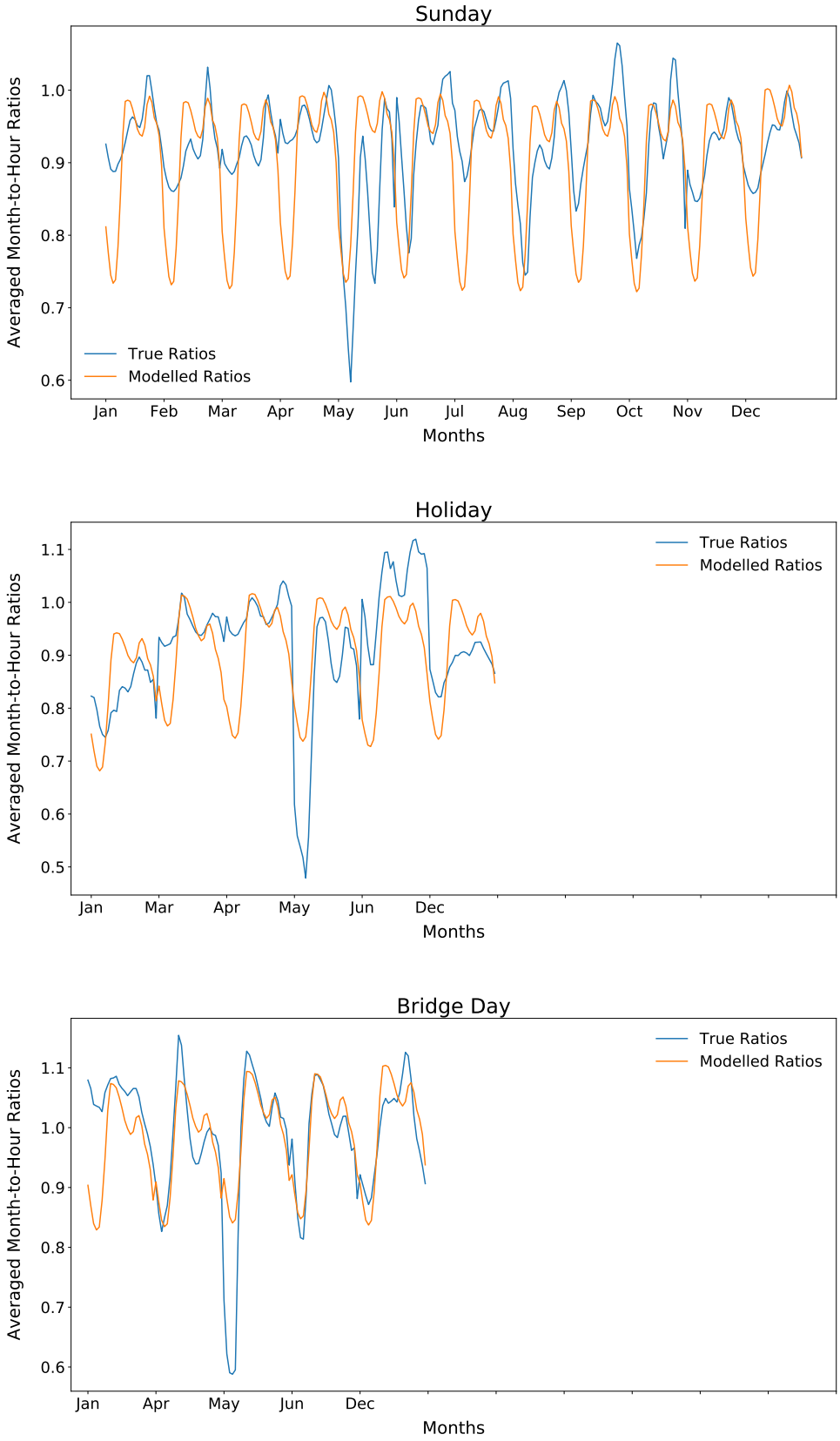


Figure A.3: Intra-yearly seasonalities for the benchmark model. Month-to-hour ratios averaged for each day-type and month for every hour of the day over the test period, starting with the hours of January and ending with the hours of December.

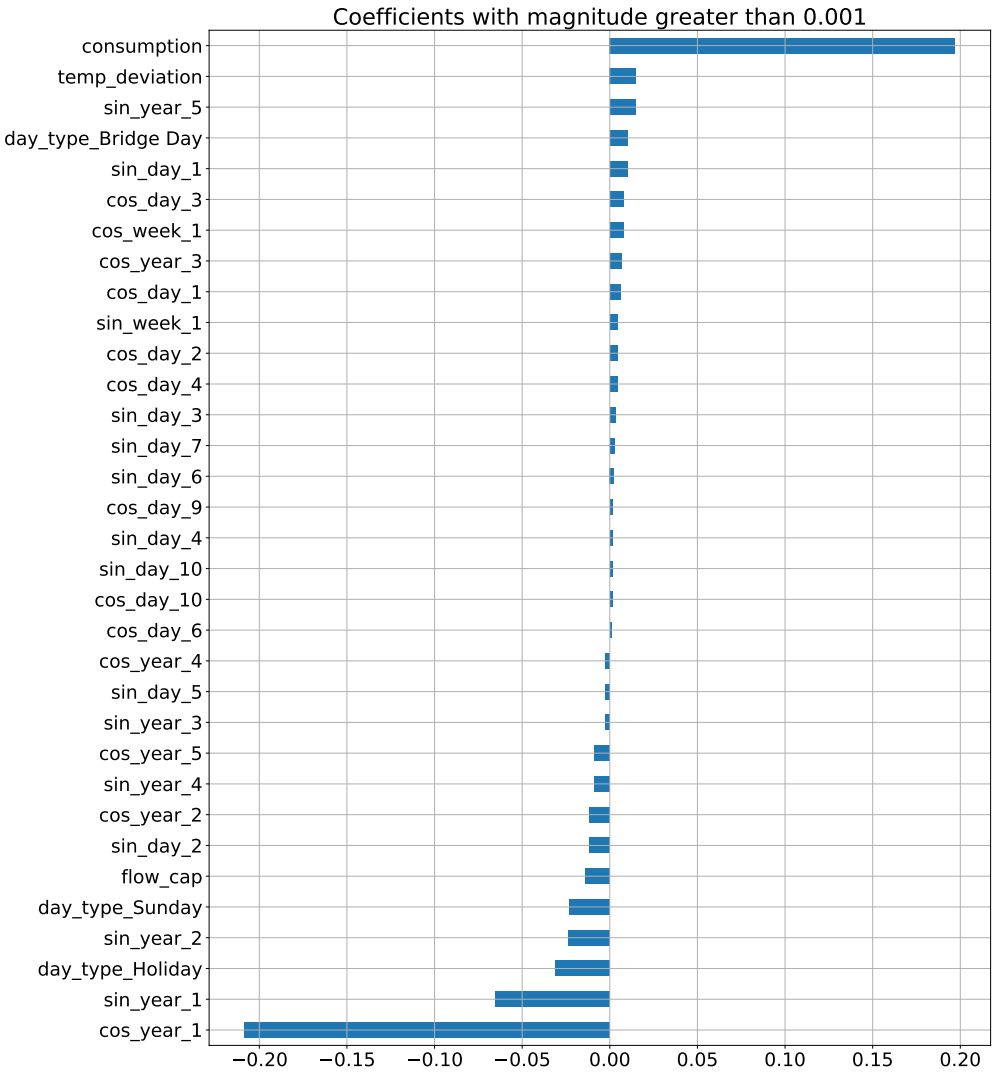


Figure A.4: Coefficients for the best-performing deterministic model.

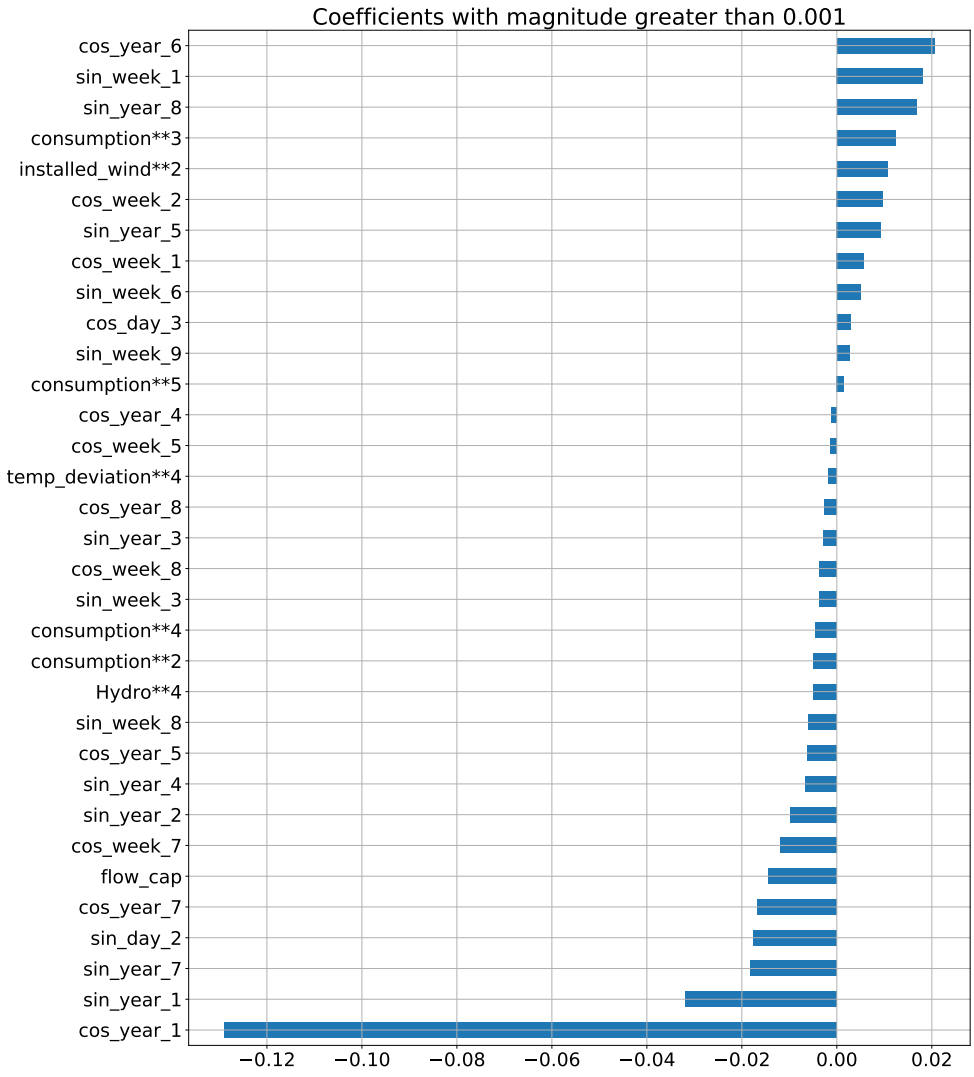


Figure A.5: Coefficients for the best-performing regularized model.

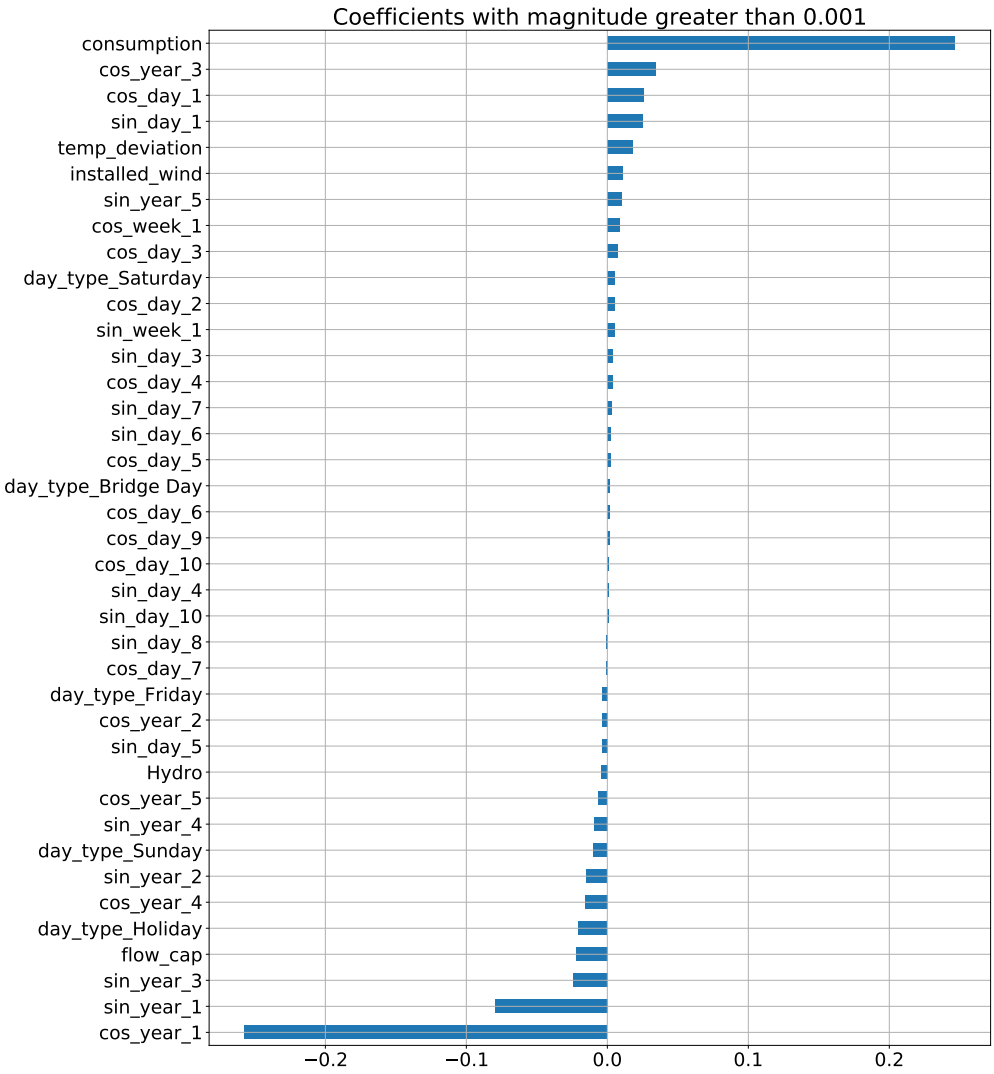


Figure A.6: Coefficients for the best-performing stochastic model.

Bibliography

Beolet, Vivian, Cyriel de Jong, and Emilian Enev (2014). *Improved hourly shaping using renewable power production*. Accessed on 2019-05-02. URL: https://www.kyos.com/wp-content/uploads/2016/06/KyCurve_with_Renewables_analysis_report_20140319.pdf.

Bishop, Christopher M. (2006). *Pattern Recognition and Machine Learning*. Cambridge CB3 0FB, United Kingdom: Springer Science+Business Media, LLC.

Brancucci Martinez-Anido, Carlo, Greg Brinkman, and Bri-Mathias Hodge (2016). "The impact of wind power on electricity prices". In: *Renewable Energy* 94, pp. 474–487. DOI: https://doi.org/10.1016/j.renene.2016.03.053.

Burger, Markus, Bernhard Graeber, and Gero Schindlmayr (2014). *Managing Energy Risk*. West Sussex, United Kingdom: John Wiley Sons, Ltd.

Energiinspektionen (2017). "The Swedish Electricity and Natural Gas Market 2016". In: Ei R2017:06. Accessed on 2019-05-04, pp. 891–921. URL: https://www.ei.se/PageFiles/310277/Ei_R2017_06.pdf.

European Parliament (2016). *Understanding electricity markets in the EU*. Accessed on 2019-05-15. URL: http://www.europarl.europa.eu/RegData/etudes/BRIE/2016/593519/EPRS_BRI(2016)593519_EN.pdf.

Green, Rikard (2014). "A Power Market Forward Curve with Hydrology Dependence. An Approach based on Artificial Neural Networks". In: *SSRN Electronic Journal*. DOI: 10.2139/ssrn.2428789.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009). *The Elements of Statistical Learning*. New York, NY, USA: Springer-Verlag New York Inc.

Hastie, Trevor, Robert Tibshirani, and Martin Wainwright (2015). *Statistical Learning with Sparsity. The Lasso and Generalizations*. New York, NY, USA: Springer Science+Business Media, Inc.

Hildmann, Marcus, Gregoire Caro, et al. (2017). "What Makes a Good Hourly Price Forward Curve?" In: *10th International Conference on the European Energy Market (EEM)*. DOI: 10.1109/EEM.2013.6607349.

Hildmann, Marcus, Florian Herzog, et al. (2011). "Robust Calculation and Parameter Estimation of the Hourly Price Forward Curve". In: *17th Power Systems Computation Conference (PSCC)*, pp. 1304–1310.

Jakobsson, A. (2016). *An Introduction to Time Series Modeling*. Lund, Sweden: Studentlitteratur.

Kaffe, Edvokia (2011). "Combined Estimation and Prediction of the Hourly Price Forward Curve". MA thesis. Zurich, Switzerland: Swiss Federal Institute of Technology (ETH) Zurich.

Kaminski, Vincent (2013). *Energy Markets*. London W1A 2HG, United Kingdom: Risk Books.

Nasdaq OMX (2019). *Bidding Areas*. Accessed on 2019-05-10. URL: https://business.nasdaq.com/trade/commodities/products/power-derivatives/nordic.html.

Nord Pool Bidding Areas (2019). *Bidding Areas*. Accessed on 2019-05-10. URL: https://www.nordpoolgroup.com/the-power-market/Bidding-areas/.

Nord Pool Power Market (2019). *The power market*. Accessed on 2019-05-12. URL: https://www.nordpoolgroup.com/the-power-market/.

Ramsay, J.O. and B.W. Silverman (2005). *Functional Data Analysis*. New York, NY, USA: Springer Science+Business Media, Inc.

Rawlings, John O., Sastry G. Pantula, and David A. Dickey (1998). *Applied Regression Analysis. A Research Tool*. Second Edition. New York, NY, USA: Springer-Verlag New York Inc.

Shiryaev, Albert N. et al. (2006). *Stochastic Finance*. New York, NY, USA: Springer Science+Business Media, Inc.

Sviland Saetherö, Audun (2017). "Hourly Price Forward Curves for Electricity Markets. Construction, Dynamics and Stochastics". PhD thesis. Duisburg, Nordrhein-Westfalen, Germany: University of Duisburg-Essen.

The European Commission (2017). *Report of the Commission Expert Group on electricity interconnection targets*. Accessed on 2019-05-04. URL: https://ec.europa.eu/energy/sites/ener/files/documents/report_of_the_commission_expert_group_on_electricity_interconnection_targets.pdf.

Weron, Rafal (2014). "Electricity price forecasting. A review of the state-of-the-art with a look into the future". In: *International Journal of Forecasting* 30 (4), pp. 1030–1081. DOI: https://doi.org/10.1016/j.ijforecast.2014.08.008.