Electricity Prices Forecasting

Asraa Amin
Supervisor: Dr. Ludger Evers

This report is submitted in partial fulfilment of the requirements for the degree of
Master of Science (MSc)

School of Mathematics & Statistics
Univerity of Glasgow

May 2020

Abstract

Much of the electricity prices forecasting has been presented for short-term forecasting horizon. However, forecasting electricity prices for longer forecasting horizons is becoming more important especially for risk management, investment planning processes and derivatives pricing.

The research presented in this dissertation lies in the area of electricity prices forecasting. In particular, we study the performance of various forecasting models when applied to electricity prices including TBATS, STLM, DR, GAM, GAMM and RF model. We focus on analysing the performance of these models for different forecasting horizons. The analyses are performed in R programming language.

We study the electricity prices forecasting for three forecasting horizons: day-ahead, week-ahead and month-ahead. We investigate the impact of calendar effects on forecasting the electricity prices. Calendar effects include hour of the day, day of the week, day of the year, month of the year and whether the day is a normal weekday, a weekend or a public holiday. Historical electricity prices is our main predictor. For each model, we use different calendar adjustments and mathematical transformations on the data to make the forecasting simpler. Then we examine the performance of each model for the three forecasting horizons. We evaluate the performance of each model using time series cross validation, rolling forecasting origin technique. We measure the forecast accuracy in terms of root mean square error (RMSE) and mean absolute error (MAE).

Our results show that the GAMM and GAM models have the best performance in day-ahead forecasting as well as week-ahead forecasting in terms of RMSE and MAE. While other models have comparable performance. As we increase the forecasting horizon, the forecasting error also increases. However, we find that the TBATS model is more robust against increasing forecasting horizons when compared to other models. For month-ahead forecasting, the TBATS model has the best forecasting accuracy compared to other models. The DR model comes after TBATS model with a slight difference in MAE and RMSE.

Then we study the statistical significance of the difference in forecasting accuracy between all considered forecasters using Diebold-Mariano test. The TBATS model proves to have statistically significantly better accuracy than all other models for month-ahead forecasting. While in week-ahead, GAM and TBATS models have statistically significantly better forecasting accuracy compared to other models. In day-ahead forecasting, the accuracy of the GAM model is statistically significantly better than all models except for the TBATS model.

Our results confirm and improve on the previous investigations of the electricity prices forecasting. In particular, we thoroughly investigate three forecasting horizons for six different models and compare them rather than just focusing on day-ahead forecasting. In addition, we investigate the differences in forecasting accuracies using the Diebold-Mariano test to confirm our results.

Acknowledgements

I would like to express my deep appreciation to my supervisor Dr Ludger Everefor his support, understanding, help and guidance throughout this project. I would also like to extend my gratitude to everyone in the department for being so helpful. I would like to thank my family for their unreserved help. My special thank you goes to my husband Khalid for providing me with unfailing support and lifting my spirit when I was down. To my parents: thanks for your love and unwavering faith throughout these years. Without you, none of this would have been possible. I am indebted to you more than I can ever repay.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

Unlike other commodities, electricity has uncommon characteristics that can affect prices considerably. First, electricity is a non-storable commodity. So we need constant balance between supply and demand to maintain system stability. Also, demand depends on the time, e.g. hour of the day, day of the week and time of the year. Thus, demand is inelastic over the short-term. In addition, external weather conditions affect the load and generation[LRS18]. Moreover, the integration of neighbouring markets is crucial [LRVS18a]. These characteristics result in non-stationary prices with high volatility and unexpected peaks.

One of the important global research areas since the foundation of the deregulated whole-sale electricity market is electricity price forecasting. All market participants such as investors, load serving bodies, traders and generating companies, require precise information about supply-demand balance as well as electricity price forecasting to maximize their profitability, set bidding strategies and define their development plans [STMY15][HTWI09]. Due to the electricity's distinct features and bidding strategies of the market participants, electricity price forecasting can be very complex and challenging. The main influencing factors on electricity prices are shown in Figure 3-1[HTWI09]

Different inputs are used for different forecasting methods to represent factors with significant effect on electricity price forecasting. Unlike electricity load time series, the electricity price series has more complex patterns including strong seasonality, various calendar effects and non-constant mean and variance. The most suitable model for electricity price forecasting should consider the stochastic components and the deterministic patterns. The stochastic components are shaped by uncertain parameters such as power plant outages. Also, they are characterised by the distinct features of the electricity energy mentioned before such as the non-stability which result in prices with high volatility and sudden peaks.

Many techniques and models have been developed for electricity price forecasting. In [HTWI09]and

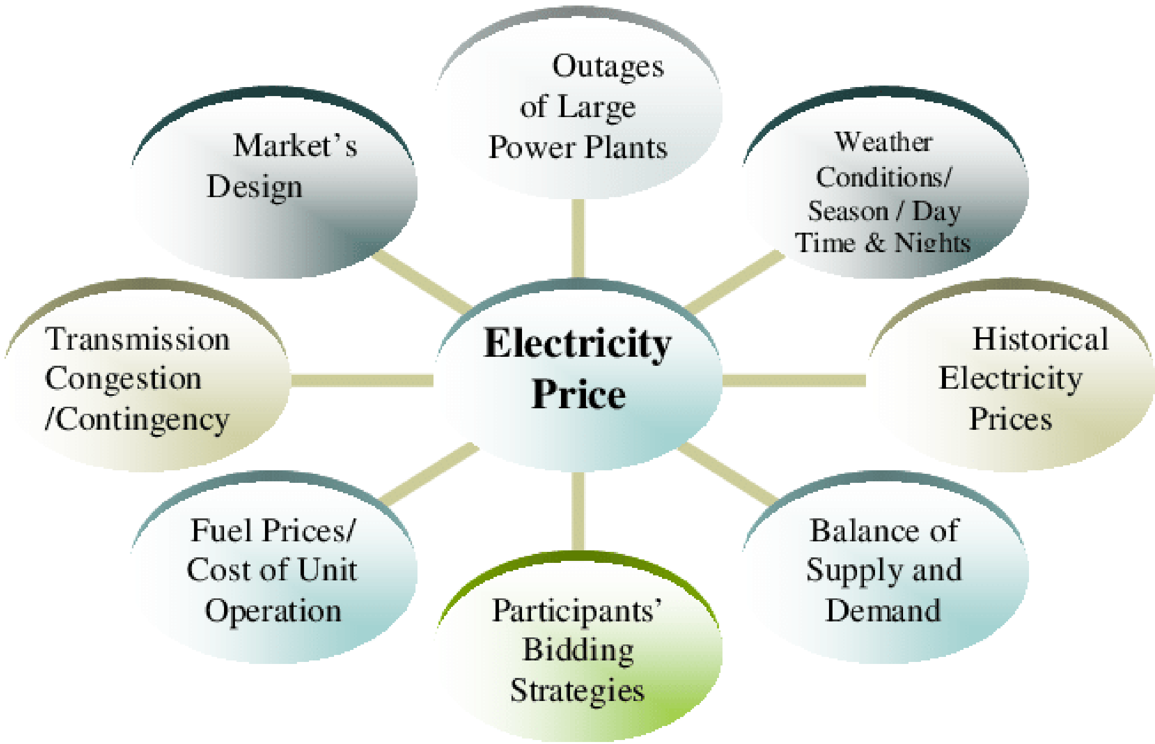


Figure 1-1: Influencing factors on electricity prices [HTWI09]

[ASK09], techniques are classified into: time series, simulations and data mining. In this thesis, time series and data mining methods are used to forecast electricity prices. Time series models use historical series of electricity prices as the main input. It shows good performance with low frequency data like weekly patterns while machine learning methods are good in predicting the nonlinear behaviour of high frequency data with high volatility, e.g. hourly data. The reasons are machine learning models don't depend on parameters to model the data and they are easier to adjust showing reliable performance when applied to complex and nonlinear data like the electricity prices.

In general, the effectiveness of each model seems to depend on many factors including the forecasting horizon. Short term electricity prices forecasting is important for day-to-day market operations which involves forecasts from several minutes to few days ahead. Medium term electricity prices forecasting, from a few days to a few months ahead, has a significant role in balance sheet calculations and risk management. While long term electricity prices forecasting, from months to years, is an important part of investment profitability analysis and planning [Wer14].

1.2 Dissertation Scope

Most of the forecasting methods have been proposed to forecast short-term electricity spot prices. In this dissertation, we study different statistical and machine learning methods to forecast electricity spot prices in UK including TBATS model, STLM model, Dynamic Harmonic Regression model

(DR), Generalised Additive Model (GAM) and Random Forest model (RF). We focus our study on the performance of these methods for different forecasting horizons including day-ahead, week-ahead and month-ahead. We utilize different criteria to evaluate the performance of these models in terms of mean absolute error (MAE) and root mean square error (RMSE). We measure the accuracy of the forecasts by using time series cross validation. Then we compute the prediction intervals (PIs) to express the uncertainty in the forecasts produced by each mode. In addition, we evaluate the models' performance by checking model residual diagnostics. Furthermore, we compare the obtained forecasting accuracy by means of hypothesis testing and discusses the results.

1.3 Dissertation Organisation

This dissertation is organized into five chapters. In Chapter 2 we give a review of various forecasting models including TBATS model, STLM model, Dynamic Harmonic Regression model (DR), Generalised Additive Model (GAM) and Random Forest model (RF). We give a detailed mathematical description of the models based on time series analysis. In Chapter 3, we give a description of the electricity prices data and the performed data processing and transformations. In addition, the model setup description for the used forecasting methods is provided . Moreover, we outline model evaluation and forecasting accuracy measures. In Chapter 4 we evaluate the performance of the used forecasting methods for different forecasting horizons. Then a comparison between these models is performed. Finally, in Chapter 5, we summarize our results and outline possible directions for future work.

Chapter 2

Forecasting Models

In this chapter, an overview of different time series forecasting models is provided. Six different models are presented: Dynamic harmonic regression model (DR), TBATS model, STLM model, Generalised Additive Models (GAMs), Generalised Additive Mixed Models (GAMMs) and Random Forest Model (RF). The mathematical description of the models based on time series analysis is provided.

2.1 DR Model

Most of the time, classical regression is not enough for describing all the interesting dynamics of a time series. Examining the Autocorrelation Function (ACF) of simple linear regression model residual, shows that it discloses additional structure in the data that the model did not capture. As an alternative, the generated correlation through lagged linear relations is introduced. The Autoregressive (AR) and Autoregressive Moving Average (ARMA) were presented in[DW52]. The ARMA model assumes that the time series is a stationary stochastic process with constant mean and variance over time. ARMA model can be reduced to AR or moving average (MA) models under certain conditions. Also, If the time series is not stationary a transformation to stationary time series by differencing must be done first. The combination of differencing with Autoregression and Moving Average model (ARMA) results in Autoregressive Integrated Moving Average (ARIMA) model or non-seasonal ARIMA model as introduced in [BJRL15a] by Box and Jenkins. So, ARIMA model is used to describe a non-stationary stochastic process. To model seasonal data, additional seasonal terms are included in ARIMA model to form seasonal ARIMA model. Seasonal ARIMA model is written :

$$ARIMA \underbrace{(p,d,q)}_{\substack{\text{Non-seasonal}\\ \text{part}}} \underbrace{(P,D,Q)_m}_{\substack{\text{Seasonal}\\ \text{part}}}$$

A detailed description of ARIMA and seasonal ARIMA models are provided in section A.1. Seasonal ARIMA model is designed to be used for shorter seasonal periods like quarterly data and monthly data with seasonal periods 4 and 12 respectively. For example, daily data can have an annual seasonality of length 365. If the seasonal period is n, there are n-1 parameters to be estimated and that is practically impossible. Therefore, for long seasonal periods, dynamic regression with Fourier terms[HA18] is better than seasonal ARIMA models. Harmonic regression method uses Fourier terms to model the seasonal pattern while ARMA error is used to model the short term time series dynamics. It allows data with one seasonal period or multiple seasonal periods and any length of seasonality. In addition, it is easier to apply to several types of time series without the need of re-identification and re-modelling when new data becomes available. However, the only drawback of dynamic harmonic regression is the assumption of fixed seasonality over time. In practice this is a valid assumption and not an issue as seasonality is usually constant except for long time series.

For multiple seasonality, Fourier terms are added for each seasonal period. The best model with minimum Akaike Information Criteria (AICc) is obtained by choosing the number of Fourier terms for each seasonal period.

2.2 TBATS Model

TBATS is a trigonometric exponential smoothing state space model with Box-Cox transformation, ARMA errors, Trend and Seasonal components. It's a combination of Fourier terms with an exponential smoothing state space model and a Box-Cox transformation, and it is proposed by De Livera, Hyndman, & Snyder in [LHS10] for modelling time series with multiple seasonality in an automated manner. In TBATS model, Box-Cox transformation is used for non-linearity and the trigonometric representation is utilized to model the seasonality. Stationary ARMA representation is used to capture any autocorrelations in the residuals. If the time series at t is y_t, $y_t^{(\lambda)}$ is the time series at time t after Box-Cox transformation. After time series transformation, decomposition results into level component l_t, growth compoenet b_t, an irregular component d_t, and seasonal component $S_t^{(i)}$ with seasonal frequencies m_i [NMI18]. These TBATS model's components are explained in details in section A.2 [LHS10].

This model has two advantages: first it uses two seed states despite the period length and second it can model seasonal part with non-integer lengths i.e. leap year with a season length 365.25. One of the drawbacks of the TBATS model is that it does not allow for adding covariates. Also, it takes a lengthy amount of time to estimate the parameters particularly for long time series. In contrast to a dynamic harmonic regression model, the TBATS model allows seasonality to change slowly over time. Also, it sometimes gives very poor result as for any automated modelling. In the estimation process, TBATS considers many models and fits the best model based on the minimum AIC. It considers modelling with and without the following options and chooses the best [LHS10]:

- Box-Cox transformation
- With and without Trend Damping
- With and without Trend
- With and without ARMA(p,q) process used to model residuals
- Non-seasonal model
- Various amounts of harmonics used to model seasonal effects

2.3 STLM model

STL is Seasonal and Trend decomposition using Loess, where Loess is a technique for estimating nonlinear relationships. The STL method was introduced by Cleveland, Cleveland, McRae, & Terpenning[CCMT90] for decomposing a time series into trend, seasonal and remainder components. While STL is beneficial for studying time series, it can also be used for forecasting. The additive time series can be described as:

$$\hat{y}_t = \hat{S}_t + \hat{A}_t \tag{2.1}$$

where $\hat{A}_t$=$\hat{T}_t + \hat{R}_t$ is the seasonally adjusted component.
The seasonal component $\hat{S}_t$ and the seasonally adjusted component $\hat{A}_t$ are forecasted separately to estimate the decomposed time series. Any non-seasonal forecasting approach can be used to forecast the seasonally adjusted component such as non-seasonal ARMA model. For the seasonal component forecasting, seasonal naive method is used by assuming the seasonal component is constant or changing very slow and taking the last year of the estimated component.

In this thesis, for the seasonally adjusted component, simple exponential smoothing with additive errors is used because it's suitable for forecasting data with no clear trend or seasonal pattern. It is described in details in section A.3. For seasonal naive method, each forecast is equal to the last observed value for the same time of the year. The forecast for time $T + h$, for the seasonal naive method, is written as [HA18]:

$$\hat{y}_{T+h|T} = y_{T+h-m(k+1)} \tag{2.2}$$

Where m is the seasonal period and k is the integer part of (h-1)/m.

2.4 GAM and GAMM Models

Generalised Additive Model (GAM) [Woo17][MSL19a] is a Generalised Linear Model (GLM) in which the response variable is modelled by predictors in the form of some smooth functions to capture the nonlinearities underlying the data. The GAM model is written as:

$$g(E(y_i)) = \beta_0 + f_1(x_{i1}) + \cdots + f_p(x_{ip}), i = 1, \cdots, N \tag{2.3}$$

Where:

y_i: is response variable coming from an exponential family distribution.

g: is a link function which can be identity,logarithmic or inverse,etc.

$x_{i1}, \cdots, x_{ip}$: are independent variables.

β_0: is an intercept.

$f_1, \cdots, f_p$: unknown non-parametric smooth functions.

The smooth function f is represented as a sum of basis functions b_j and their regression coefficients β_j as follows:

$$f(x) = \sum_{j=1}^{q} b_j(x)\beta_j \tag{2.4}$$

Where:

q: basis dimension.

More detailed description and determination of estimates is given in section A.4.

In GAMs [MSL19b][Woo17],there are different choices of smoothing splines such as cubic regression splines, cyclic regression splines, thin plate regression splines, P-splines, etc. The choice of the level of rank and order, the number and location of the knots and the number of predictors, make each model different from one another. Also, it might be necessary to add predictors interactions including basic forms like multiplication and more complex form like tensor products [MSL19b]. For double and multiple seasonal time series, interactions are very important and can be included to the model in four different ways :

- Multiplications of two predictors : $x_1 \times x_2$.
- Use smooth function to one predictor $f_1(x_1) \times x_2$.
- Use smooth function for both predictors $f_1(x_1) \times f_2(x_2)$ or $f(x_1, x_2)$.
- Use tensor product interactions $f_1(x_1) \otimes f_2(x_2)$.

GAMs models assume the residuals (errors) are i.i.d. For time series regression, this assumption is not valid. The present time series values are strongly correlated with the past values which is called autocorrelation. So residuals of the model will also be correlated. This autocorrelation might result in a biased estimation of the residuals standard deviation and thus wrong calculated confidence intervals and p-values. By adding the AR model for errors in our GAM model:

$$y_i = x_i^T \beta + \epsilon_i, \epsilon_i = \phi\epsilon_{i-1} + \upsilon_i \tag{2.5}$$

This equation represents a classical AR(1) model with ϕ being an estimated autoregressive coefficient. MA models also can be added to have ARMA(p,q) models with different values for p and q [BJRL15b]. The result of adding the ARMA model for errors in the GAM model is a GAMM model. More detailed description of ARMA models is given in section A.1.

2.5 RF Model

Decision trees suffer from high variance. Bootstrap aggregation, or bagging, is a technique used to reduce the variance of statistical learning method. By taking repeated samples from the training data set, separate prediction model (decision trees) are built for each sample and the resulting predictions is averaged. This process reduces the variance and increases the accuracy. Train the model on the $b^t h$ bootstrapped sample to get $\hat{f}^{*b}(x)$ and average all the predictions resulting in :

$$\hat{f_{bag}} = \frac{1}{B}\sum_{b=1}^{B}\hat{f}^{*b}(x) \tag{2.6}$$

RF [HTF09] have the same concept as bagging, but it decorrelates trees. In bagging, a number of decision trees are built on a bootstrapped sample from the training set. But in RF, when building decision trees, for each split, a random sample of size m of the predictors is chosen from the total number of predictors p. The number of predictors at each split is usually chosen to be the square root of the total number of predictors. So if there is one very strong predictor in the data, on average $(p - m/p)$ of the splits will not include the strong predictor. This process decorrelates the tress and makes the resulting trees more consistent and reliable. If B is increased, random forests as well as bagging will not overfit. However, random forest can still result in overfitting when the chosen trees are too deep.

Using Random forests for time series forecasting is not straightforward as it doesn't have awareness of time. It assumes the data are i.i.d, but this assumption is not valid as time series is characterized by serial dependence. Also, random forests or all decision trees based techniques are unable to predict trends. In order to use decision trees based models to forecast a univariate time series, several pre-processing steps must be done, namely as below:

- Data transformation (log transform, Box-Cox transform, etc.)
- Detrending by differencing, STL, etc.
- Time Delay Embedding
- Feature engineering(lags, Fourier transform, etc.)

Chapter 3

Model Setup

In this chapter we start by describing the electricity prices data. We then proceed to explain the performed data processing and transformations. This is followed by model setup description for various forecasting methods used including: TBATS, STLM, DR, GAM, GAMM and RF. We finally outline model evaluation and forecasting accuracy measures.

3.1 Data Processing

The electricity prices on the wholesale electricity market in the UK were collected on an hourly basis from January 2017 to January 2020. The electricity prices series contains 507 missing values in 2017 (21 days data) as shown in Figure 3-1. Some models allow for missing values and can deal with them without causing errors such as the R functions for ARIMA models and DR models. However, other modelling functions can't handle missing values including TBATS and STLM models. In such cases, the missing values are replaced with estimates as seen in Figure 3-1. These estimates are calculated using simple linear interpolation for non-seasonal data. While for seasonal data, STL decomposition is calculated, the linear interpolation is applied to the seasonally adjusted series and the seasonal component is added back [MBB17].

As mentioned before, electricity prices series is non-stationary with high volatility and unexpected peaks. To make the time series easier to forecast with statistical models, the data is processed using a Box-Cox transformation. For each forecasting model, different inputs are used. Our response variable is the electricity prices while the main predictor variable is the past prices. Other predictor variables including calendar effects such as hour of the day, day of the week, week of the year, month of the year, the four seasons and whether the day is a normal weekday, weekend, or public holiday. In Figure3-1, the correlation of the prices with the previous values is plotted to show the autocorrelation function or ACF. The prices time series has a trend, because the ACF has positive values that slowly decrease as the lags increase. It is also seasonal, because the ACF is larger at

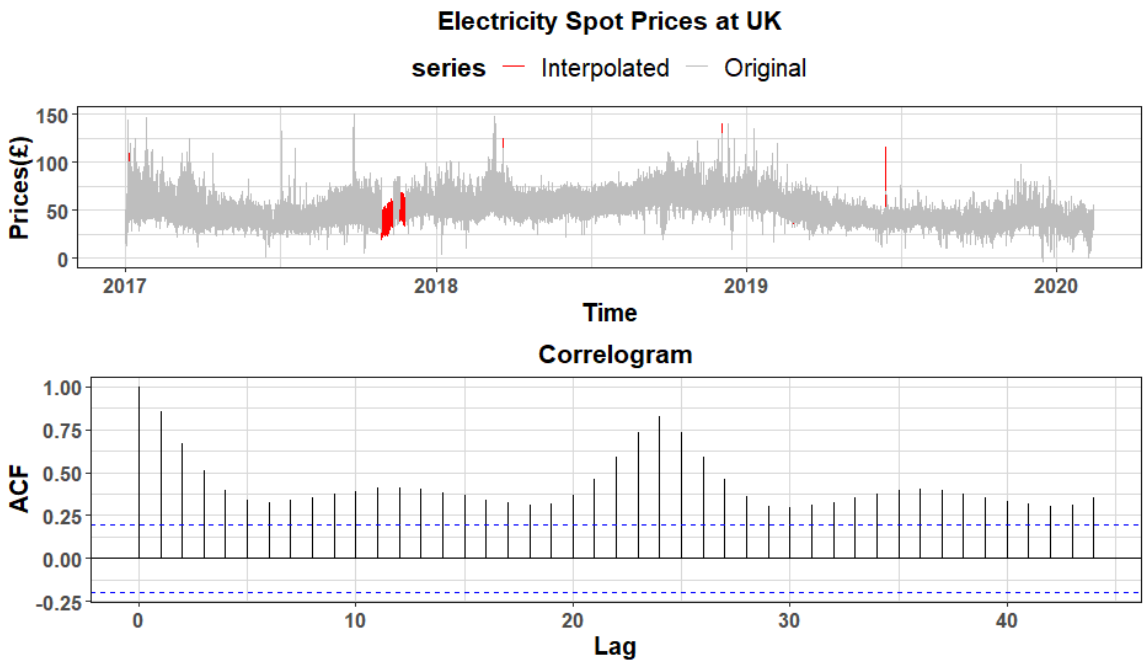


Figure 3-1: Electricity spot prices at UK

seasonal lags and multiple of the seasonal periods than for other lags. Therefore, prices time series has a combination of trend and seasonality.

Electricity price series has three types of seasonality: daily pattern, weekly pattern and annual pattern with 24, 168 and 8766 seasonal periods respectively. The goal of this study is to produce short term forecasts for one month, one week and one day-ahead within year 2019. Figure 3-2 shows the difference between weekends and weekdays in the average daily electricity price trend for four selected months in 2018. The price is relatively low for the first hours of the day before it goes up to its first peak around 9am followed by a reduction to the minimum around 3pm, rising again to another peak around 7pm and followed again by a reduction to the night level. Moreover, the figure shows a weekly pattern, as weekend spot prices are constantly above weekdays early in the first hours of the day and the last hours and almost below weekdays in-between. During spring and summer time, the slope is steeper mid-day, while in fall and winter the price declines more permanently.

3.2 Model Setup

3.2.1 TBATS Model

TBATS model is applied to the log transformation of the training data. First, the electricity prices series is decomposed into its components as mentioned before in section 2.2. The randomness is removed from the seasonal components by the trigonometric formulation without destroying the effective parts. The obtained decomposition from the TBATS model is represented in five parts in

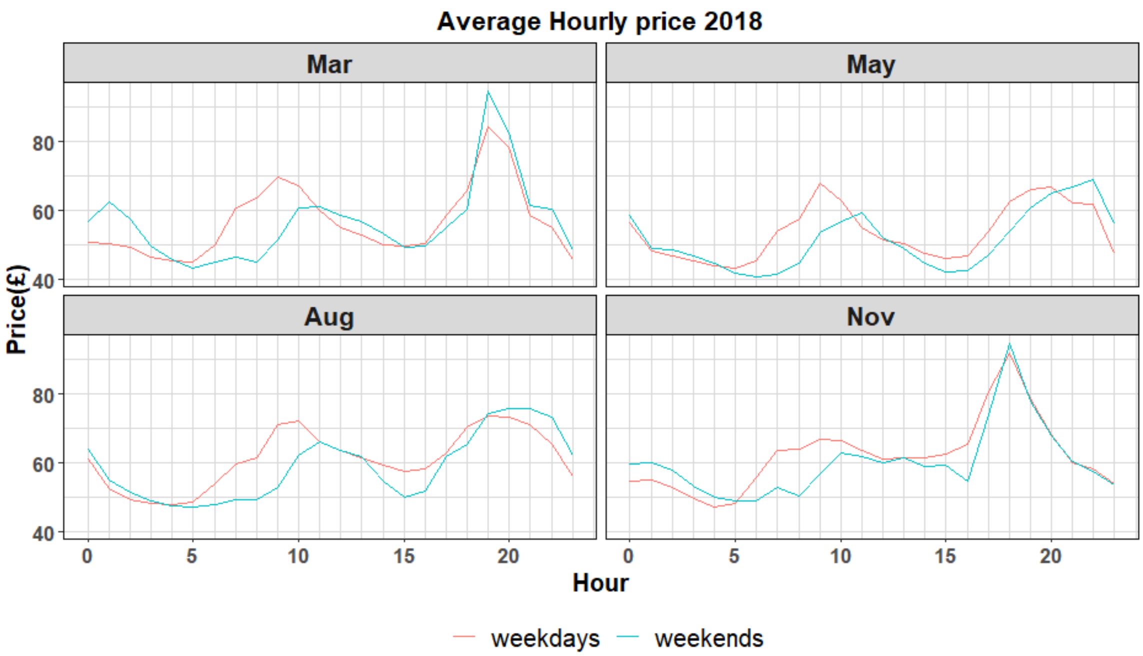


Figure 3-2: Average hourly price for selected months in 2018

Figure3-3. The first part represents the observed data and the second part shows the trend. The rest of the three parts display seasonality in electricity prices data, which are daily, weekly and yearly seasonality. It can be noticed that the three seasonal parts stay stable over time. Second, TBATS model parameters are estimated using root mean square error. Last, the appropriate ARMA orders are selected and the best model is chosen using the AIC. The obtained TBATS(1,5,1, -,<24,1>,<168,1>,<8766,1>) model represents:

- No Box-Cox transformation ($\lambda = 1$).
- The order of ARMA error are (5,1).
- No damping trend.
- The number of harmonics are $k_1 = 7$, $k_2 = 6$ and $k_3 = 4$.
- Seasonal periods are 24,168 and 8766 for daily seasonality (24 hours a day), weekly seasonality (24 hours a day and 7 days a week) and yearly pattern (24 hours a day and 365.25 days a year) respectively.
- The total number of initial seasonal values is 6.

3.2.2 STLM Model

STLM model is applied to the training data and log transformation is used by choosing the Box-Cox transformation parameter $\lambda = 0$. As mentioned before in section 2.3, the electricity prices

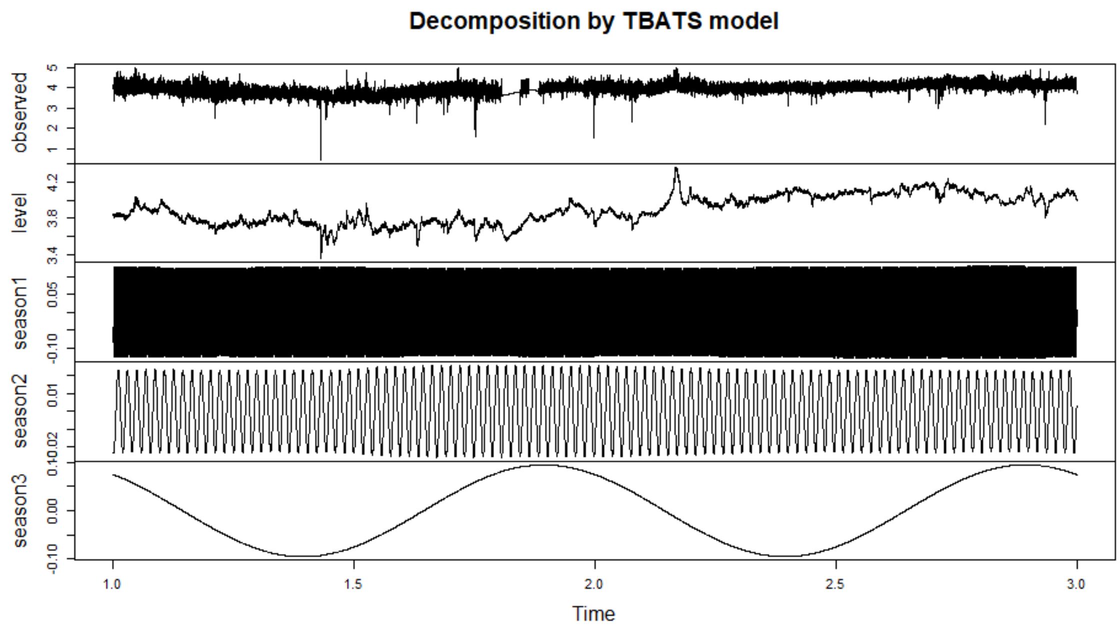


Figure 3-3: Trigonometric decomposition of electricity prices data

series is decomposed into its components. The obtained decomposition is displayed in six panels in 3-4. The first panel is the observed data. The second panel is the trend component. The next three panels are the seasonal components, which are daily, weekly and yearly seasonality respectively. Removing the trend and the seasonal components results in the error component that is presented in the last panel. Looking at the vertical scales of the STL decomposition plot, the trend and weekly seasonality have narrow ranges in comparison to the other components as the weekly seasonality is weak and the trend is hardly seen in the data. When forecasting the seasonally adjusted component, the initial values for the ETS model and the smoothing parameter α are estimated by minimizing the sum of squared errors (SSE). The estimated value of α is 0.8854 and the estimated initial states ℓ_0 is 3.9799.

3.2.3 DR Model

The DR model is applied to the training data and the log transformation is taken by choosing the Box-Cox transformation parameter $\lambda = 0$. As mentioned before in section 2.1 , for time series with multiple seasonality Fourier terms are added for each seasonal period (24,168,8766). The number of Fourier pairs for each seasonal period with minimum AICc (best model) are found to be 5,1 and 1 respectively. To find the best ARIMA model, a combination of unit root test, minimization of the AICc and maximum likelihood estimation (MLE) is considered. The best ARIMA model for the errors is found to be ARIMA(5,1,0). The total number of degrees of freedom is 19 (the other five coming from ARMA parameters).

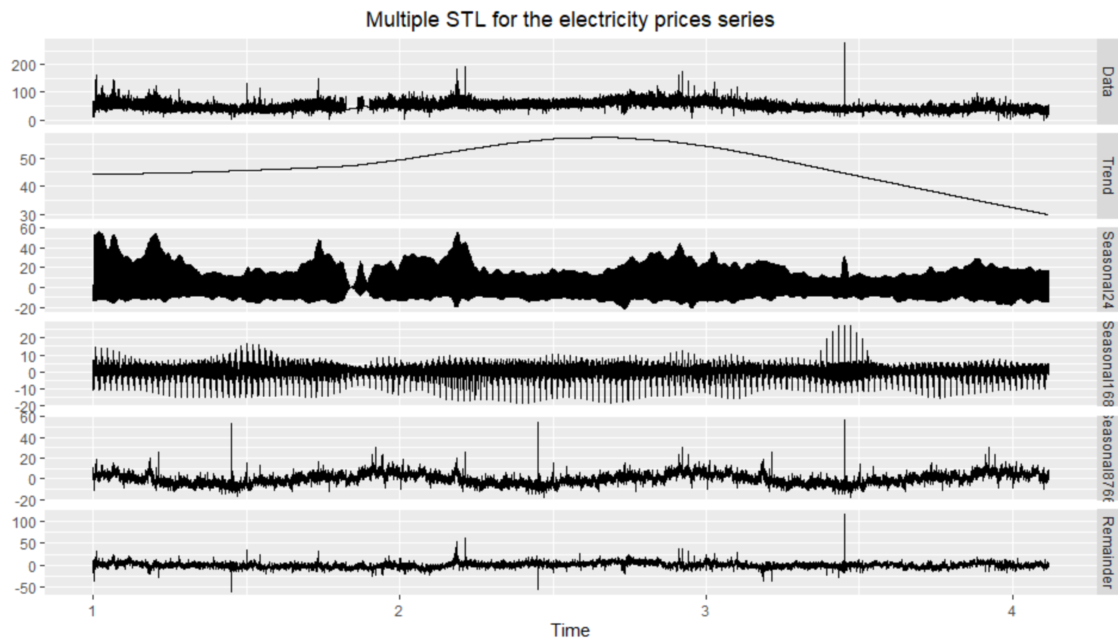


Figure 3-4: Multiple STL of electricity prices data

3.2.4 GAM and GAMM Models

GAM methods are fitted using the **mgcv** package in R presented in [Woo17]. Box-Cox transformations is applied to the training data. The function *gam()* is used for models without taking into account residuals autocorrelation. The function *gamm()* allows to specify the correlation structure for the residuals. But for large number of observations this process is very slow and memory intensive due to the high dimension of the covariance matrix of the residuals ($numberofobservationsXnumberofobseravtions$). So, instead of using *gamm()* function, the model is fitted in two stages: first stage the GAM model is applied by using *gam()* function and second stage by fitting ARIMA model to the GAM model residuals. Any type of time series model for the second stage could be used.

For the GAM model, we fit six models with different predictors and smooth functions. To compare models, fitting quality indicators are estimated such as Generalized Cross Validation (GCV) score, adjusted R^2 and percentage deviance explained. To compare models with different variables, GCV score is a good indicator,the smaller the GCV score the better the model.

The first model M_1 includes daily, weekly and yearly patterns. The individual effects of each pattern is modelled additively by univariate smoothing function. Therefore, the daily effect is assumed to be the same throughout the year and for all types of days, weekdays, weekends and public holidays. M_1 can be written as:

$$M_1 \ : \ y_t \ = \ f_1(\textit{Hour of the Day})_t + f_2(\textit{Day of the Week})_t + f_3(\textit{Month of the Year})_t + \beta_0 + \epsilon_t$$

f_i can be cubic regression splines or cubic cyclic splines. They are the same except that cubic cyclic splines has the value and the first two derivatives matching at the start and end of each time period. So, cubic cyclic splines is applied. The number of knots are set to the number of unique values in each predictor. This model assumes the three predictors are independent. This assumption is not realistic, since in Figure 3-2, it can be noticed that the effects are dependent. To account for the interactions between the three predictors, one smoothing function (one smoothing parameter λ) is used in M_2[Woo17].

$$M_2 : y_t = f_4(\textit{Hour of the Day, Day of the Week, Month of the Year})_t + \beta_0 + \epsilon_t$$

f_4 is thin plate regression spline (TPRS) [MSL19a] with one smoothing parameter λ used for all predictors. The Time of the training data is added to model M_2 to represent a long term non-linear trend. While Day of the Year is added to represent the annual cycle.

$$M_3 : M_2 + f_5(\textit{Time})_t + f_6(\textit{Day of the Year})_t$$

For f_5, cubic regression spline is used and f_6 is a cubic cyclic splines, because it would not be suitable to have discontinuity at the year ends. Two categorical variables are introduced to improve the model performance. First, type of the day with three levels week day, weekend and public holiday. Second, day of the week with seven levels.

$$M_4 : y_t = M_3 + \textit{Type of the Day} + \textit{Day}$$

The fifth model is the combination of Model M_4, the individual effect of each predictor in M_1 and the interactions between them.

$$M_5 : y_t = M_4 + f_1(\textit{Hour of the Day})_t + f_3(\textit{Month of the Year})_t$$

The last model is the same as model M_5 except that the non-linear trend is substituted with a linear trend.

$$M_6 : y_t = M_2 + \textit{Time} + f_6(\textit{Day of the Year})_t + f_1(\textit{Hour of the Day})_t + f_3(\textit{Month of the Year})_t + \textit{Type of the Day} + \textit{Day}$$

The comparison of the models is shown in Table 4.1a. It can be observed that M_5 is the best model with the highest adjusted R^2 and lowest GCV score and AIC.

Figure3-5 contains plots that show the individual effect of each predictor, which referred to partial dependence plots. The first two plots in the upper panel display the estimated annual cycle and the estimated long term trend. The second two plots in the lower panel present the daily and yearly

Model	R^2	Deviance Explained%	GCV Score	AIC
M_1	0.374	37.6	0.28155	27517.63
M_2	0.442	44.5	0.25235	25598.47
M_3	0.729	73.1	0.21822	12937.71
M_4	0.73	73.2	0.12229	12905.60
M_5	0.743	74.5	0.11673	12089.80
M_6	0.546	55	0.20561	22009.27

Table 3.1: GAM model selection measures

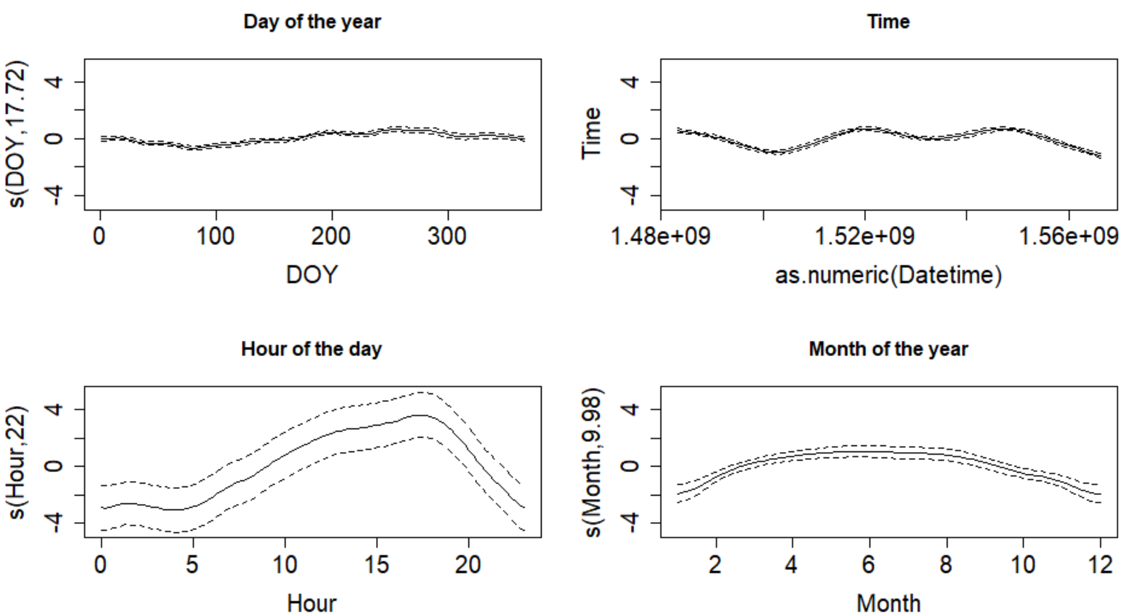


Figure 3-5: GAM model partial dependence plots.

patterns. Also, Figure C-1, Figure C-1 and Figure C-1 show the interactions between hourly, weekly and monthly. Obviously, the effects of the predictors and their interactions can be noticed. It can be observed how the prices are different between summer and winter months, the daily peaks and difference between weekdays and weekends prices.

As mentioned before, for GAMM model we fitted ARIMA model to the GAM model error. The best ARIMA model for the errors is found to be ARIMA(5,0,2)(1,1,20)[24].

3.2.5 RF Model

Using tree-based model to forecast a time series with trend can result in poor forecasts. The reason for that is decision trees are making predictions by learning decision rules from the training data. Thus, they are incapable of extrapolating data significantly different from the training data and they don't adapt well to sudden changes. RF models are facing the same problem as they are averages of decision trees. Therefore, to apply RF model, the training data series is first detrended

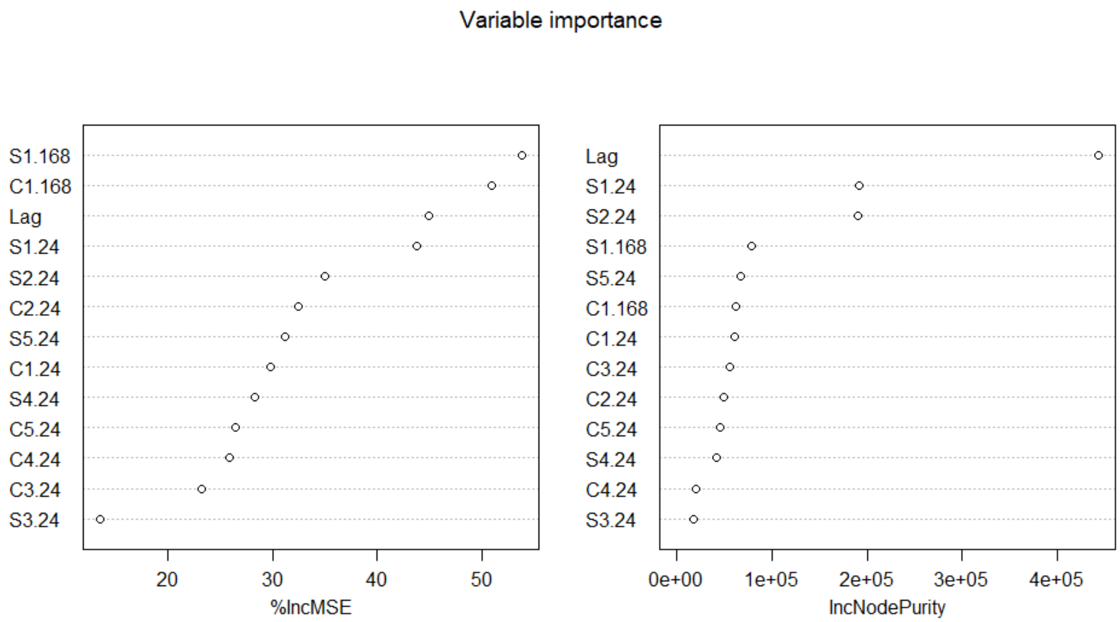


Figure 3-6: Random Forest model

by STL decomposition. The trend part is forecasted by ARIMA model. To model the seasonal component with nonlinear model like RF, it is transformed to Fourier terms, because they are more effective to nonlinear methods. The number of Fourier pairs for each seasonal period with minimum AICc (best model) are 5 and 1. A lag of the time series is added as lag feature to our model. The trend is removed first to ensure stability and the lag of the seasonal part from STL decomposition is added. Therefore, to forecast the seasonal and the remainder parts, the RF model is applied to the training data which comprises of lagged electricity prices and double-seasonal Fourier terms.

Parameters of the RF model are specified: number of trees is 200 and number of variables used in each split m is 4. Figure 3-6 shows the importance of the variables that are used as predictors. We can see that S1.168 (weekly seasonal feature in sines form), C1.168 (weekly seasonal feature in cosines form) and lag features have the best score in terms of Increase mean square error (IncMSE). It indicates that lagged values of electricity prices and weekly seasonality are the most important predictors for our model. While increase in node purity (IncNodePurity), which is calculated based on the reduction in sum of squared errors, indicates that lag feature and daily seasonal feature are the most important. The relation between the number of trees used and the model error is plotted in Figure3-7. From the plot, the optimum number of trees to be used is almost 200.

3.3 Model Evaluation

The goal of this study is to produce short term forecasts for one month, one week and one day-ahead within year 2019. The time series data $y_1, \ldots, y_T$ is split to training data $y_1, \ldots, y_N$

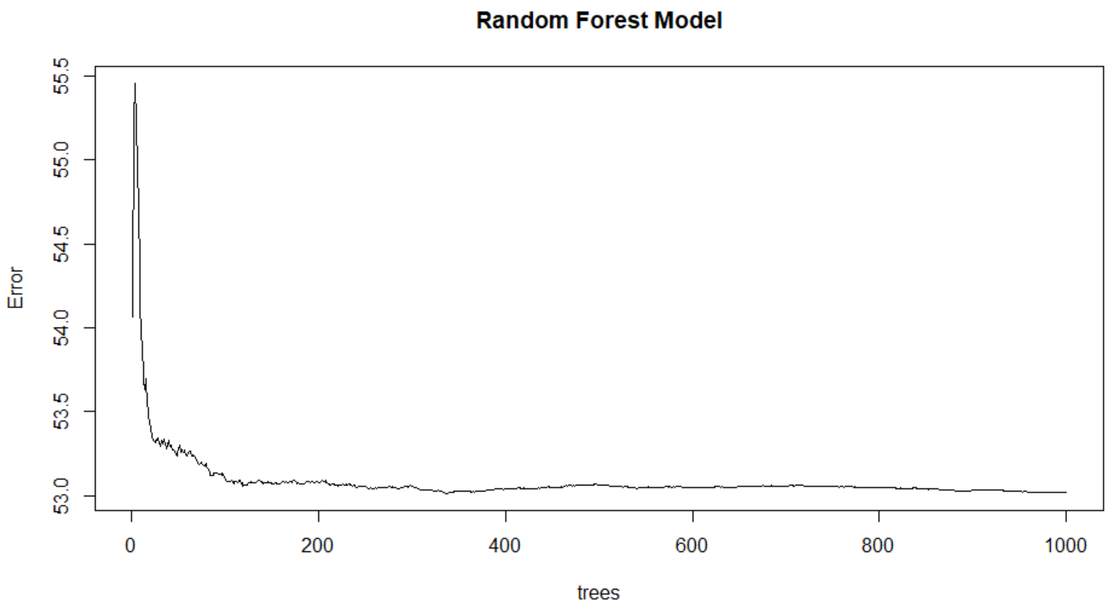


Figure 3-7: Random Forest model

to estimate models and testing data $y_{N+1}, \ldots, y_T$ to evaluate model performance. The size of the training data is two years (2017-2018) and the testing data is one year (2019).

To evaluate models, the trade-off between the model complexity and accuracy must be considered. Akaike Information Criterion (AIC) [21] and Bayes Information Criterion (BIC)[22] are in-sample (training data) accuracy measures that penalize for model complexity. To check the model accuracy, the model parameters are estimated using the training data, forecasting $T - N$ steps ahead and compared to the testing data. The forecast errors are the difference between the testing data values and the forecasts produced by using the training data set as defined in (B.1).

Different forecasting accuracy measures are defined with equations (B.4) in section B.1 and computed for each model including mean error (ME), root mean squared error (RMSE), mean absolute error (MAE), mean percentage error(MPE) and mean absolute percentage error (MAPE). The MAE and the RMSE are scale-dependent errors relative to the data scale, while MAPE is scale-independent and can be used to compare performance between different data sets. For electricity price forecasting, MAPE is problematic because it is undefined or inifinte for series values equal to zero [Mak93]. Also, it has extreme values for series values y_t close to zero. Moreover, it puts heavier penalty on negative errors than on positive errors [HA18]. Instead of MAPE, symmetric mean absolute percentage error (sMAPE) is considered and defined in section B.1, equation (B.3). Also, it's not a measure of absolute percentage error as it can also be negative. Instead of using percentage errors, Hyndman & Koehler [HK06] proposed scaled error to compare forecast accuracy on different scales, called mean absolute scaled error (MASE). The scaled error and MASE are defined in section B.1 with equations (B.9), (B.10) and (B.11). As all the forecasts are on the same scale, the MAE and RMSE are used to measure the forecasting accuracy.

3.3.1 Diebold-Mariano Test

When comparing models forecasting accuracy, higher model accuracy is not enough to ensure that the model is better. Instead, the difference in accuracy must be significant. In particular, Diebold-Mariano (DM) test [DM02] is the statistical test used to ensure the forecasting accuracy of the model is statistically significant. The DM test is applied for each hour in the forecasting horizon following the procedure in [LRVS18b]. We applied the DM test for each hour and for model pairs M_1,M_2 following the steps in section B.2 [LRS18].

3.3.2 Time Series Cross-Validation

For time series cross-validation,there are many techniques including sliding window and evaluation on a rolling forecasting origin. The rolling forecasting origin method is used. If the total number of observations is T, k observations are the minimum number of training data to produce a good model that produces h-step a head forecast, then a rolling forecasting origin works as follows[Hyn14]:

- To estimate the model parameters,observations $1, 2, \ldots, k + i - 1$ are used. For the testing data,observations $k+h+i-1$ are used. Compute the h-step forecasting error for time $k+h+i-1$.
- For total number of observations T,repeat the above step for $i = 1, 2, \ldots, T - k - h + 1$.
- Calculate forecasting accuracy measures on each test set and the average is produced across all test sets.

Chapter 4

Forecasting Results

In this chapter, we start by giving the forecasting accuracy measures of all models. We then evaluate the performance of the considered models. In addition, we assess the statistical significance of the difference in forecasting accuracy between all considered models using DM test. This is followed by a discussion of the results. We focus on two forecasting accuracy measures: mean absolute error (MAE) and root mean square error (RMSE). We evaluate the models under each one of them by checking model residuals. We analyse and discuss the results of each model. Finally we compare the results of all models.

4.1 Main Results

After processing the data and model setup description, the forecasting accuracy measures are computed on the test data set for all models. They are obtained for three forecasting horizons: month-ahead, week-ahead and day ahead, which are represented in Tables 4.1a,4.1b and 4.1c respectively. The computed forecasting accuracy measures include sMAPE, MASE, MAE and RMSE. We focus on MAE and RMSE as forecasting accuracy measures to compare models. Figure4-2 shows boxplots of the RMSE and MAE distributions for various models. The results are split by different forecasting horizon. For each model, the boxplot shows the median as well as 25 and 75% quantiles of the RMSE and MAE. In addition, the plot shows the mean as 'darkred' dots. From these results we have three key insights:

- For month-ahead forecasting the TBATS model has the lowest RMSE and MAE, while the GAM model has the highest RMSE and MAE.
- In contrast, the TBATS model has the highest RMSE and MAE for week-ahead forecasting while the GAMM has the lowest.
- Similarly, the GAMM model has the lowest RMSE and MAE for day-ahead forecasting while

the DR model have the highest.

Also, Figure 4-1 presents examples of the electricity prices forecasts from various models against the real values for the three forecasting horizons mention before. It can be noticed that it is hard to distinguish between several models forecasts and the real values of prices except for some prices peaks. More detailed results have been obtained for different models and will be discussed in the following sections.

Model	sMAPE[%]	RMSE	MAE	MASE
STLM	16.97	9.28	7.35	1.72
TBATS	**14.66**	**7.89**	**6.13**	**1.46**
DR	16.7	8.70	6.76	1.7
GAM	18.4	9.63	7.83	2.27
GAMM	17.8	9.3	7.53	2.1
RF	17.89	9.56	9.63	1.8

(a) Month-ahead.

Model	sMAPE[%]	RMSE	MAE	MASE
STLM	15.2	7.20	5.87	1.43
TBATS	16.83	7.8	6.39	1.57
DR	15.3	7.36	5.74	1.43
GAM	11.4	5.82	4.37	1.42
GAMM	**11.7**	**5.5**	**4.31**	**1.21**
RF	14.61	7.06	5.54	1.34

(b) Week-ahead.

Model	sMAPE[%]	RMSE	MAE	MASE
STLM	14.16	5.90	4.91	1.18
TBATS	15.2	6.28	5.32	1.23
DR	15.02	6.59	5.35	1.24
GAM	12.82	5.44	4.37	1.49
GAMM	**11.66**	**4.91**	**4.01**	**1.2**
RF	15.1	6.41	5.31	1.23

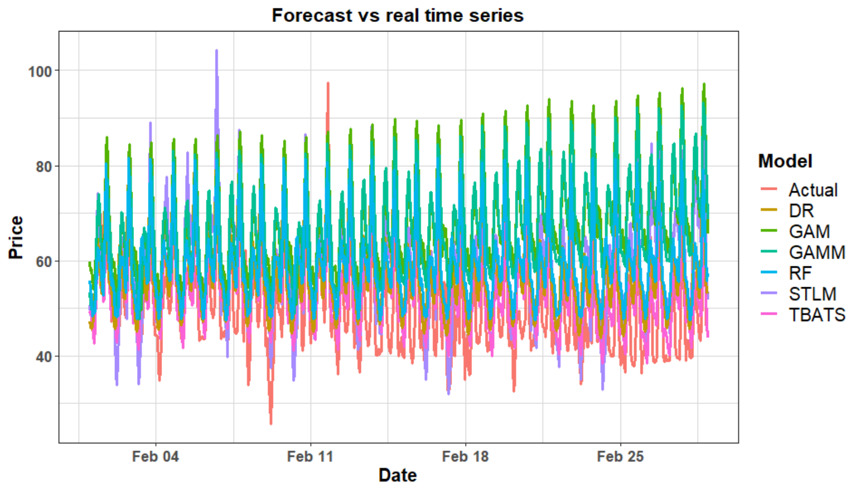
(c) Day-ahead.

Table 4.1: Accuracy measures for different forecasting horizons.

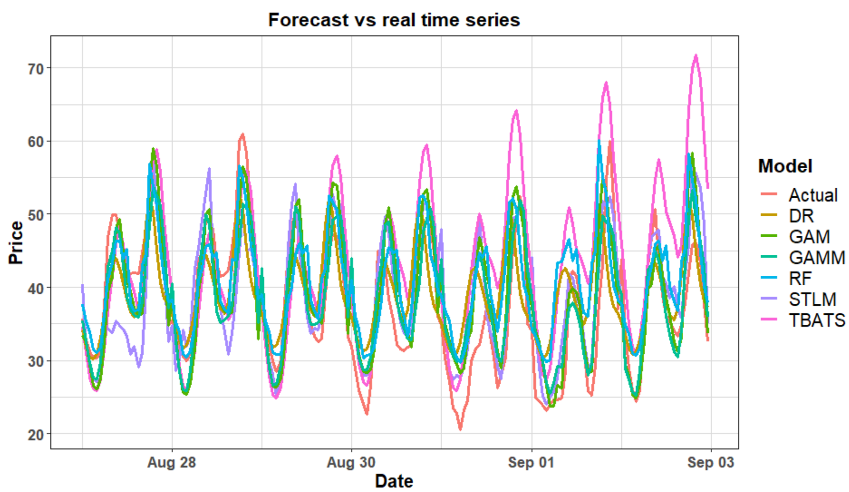
4.1.1 TBATS Model

The process of fitting the TBATS model is entirely automated. For long time series such as electricity prices, the TBATS model is slow to estimate. From the month-ahead forecasting as seen in Table4.1a, the TBATS model has significantly better precision than all other forecasting methods with RMSE 7.89 and MAE 6.13. Figure4-2 below shows how the TBATS model has the highest MAE and RMSE compared to other models for shorter forecasting horizon like week-ahead forecasting. It can be noticed that the TBATS model is more robust against increasing forecasting horizons than other models. This is because the TBATS models allow seasonality to change slowly over time. Also, the trigonometric terms in the TBATS model can deal with multiple non-integer seasonality. Our results are consistent with the those published in [VGT+14] and [NMI18].

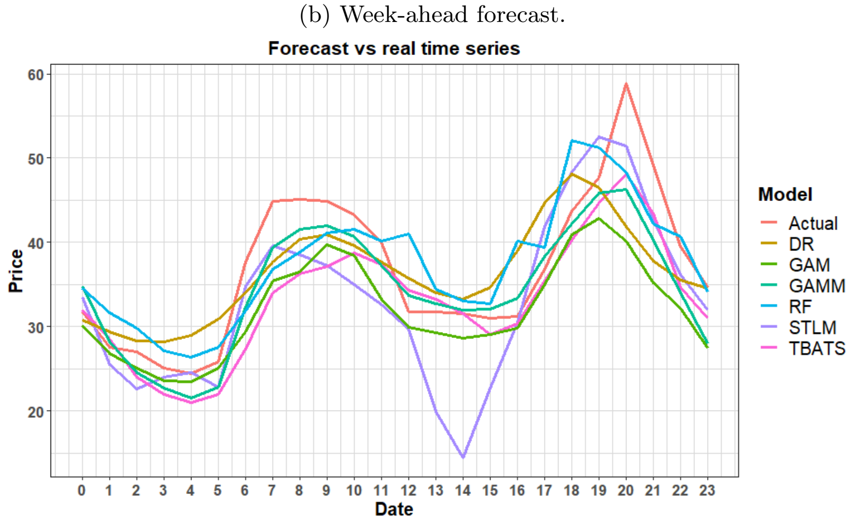
To express the uncertainty in the forecasts, we calculated the prediction interval (PI) within which the value of y_t is expected to be located with certain probability. The prediction interval for week-ahead forecasting from the TBATS model is depicted with a shaded region in Figure 4-3a. Although the TBATS model has wider prediction interval most of the times, it can be noticed that



(a) Month-ahead forecast.

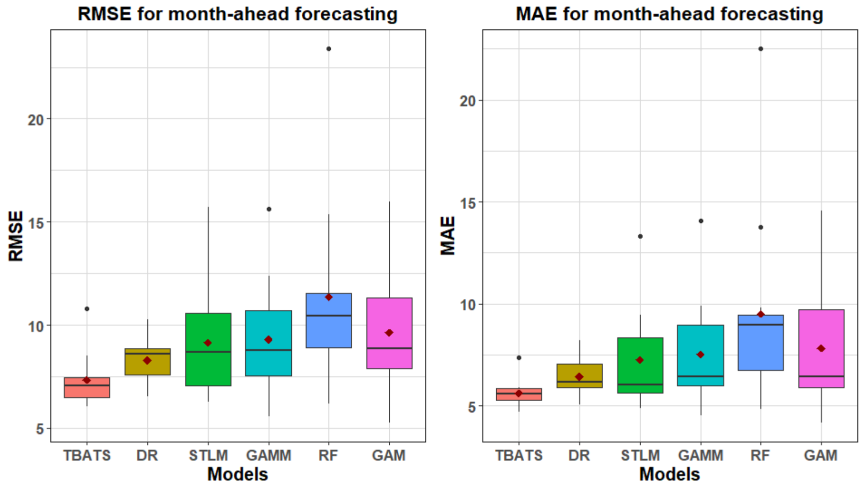


(b) Week-ahead forecast.

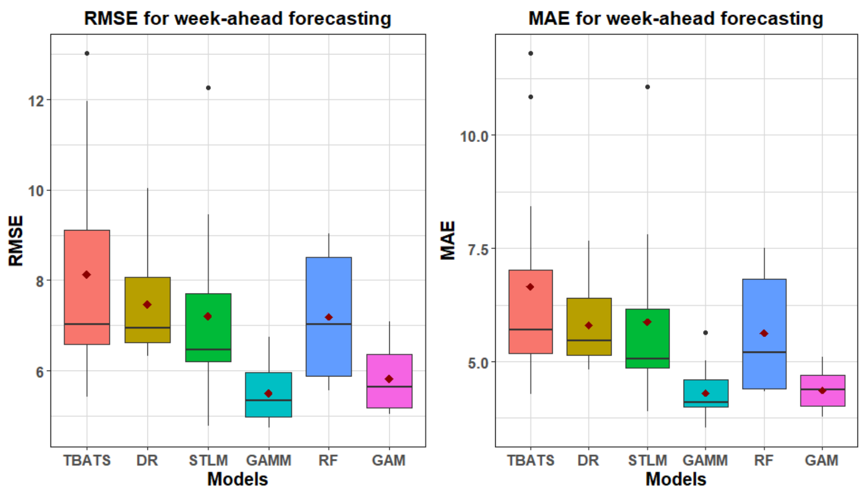


(c) Day-ahead forecast.

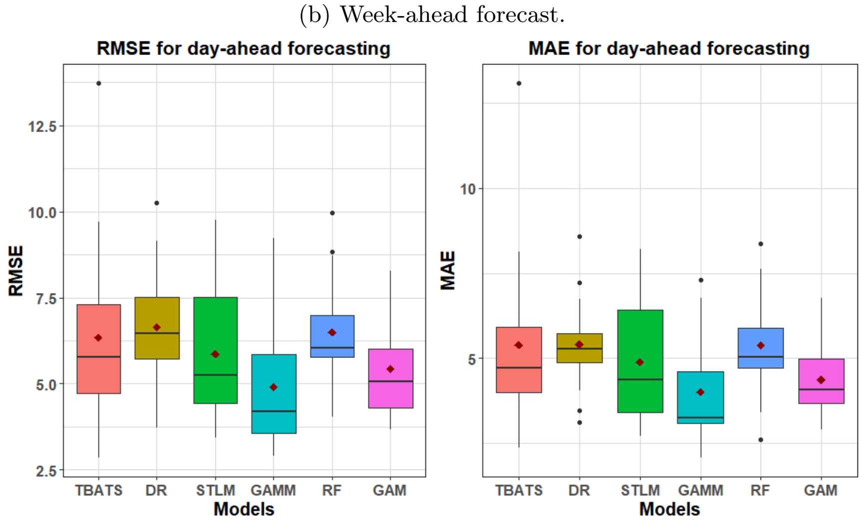
Figure 4-1: Example of monthly, weekly and daily forecast for all models.



(a) Month-ahead forecast.



(b) Week-ahead forecast.



(c) Day-ahead forecast.

Figure 4-2: RMSE and MAE for all models.

Model	Estimated σ of residual
STLM	0.055
TBATS	0.112
DR	0.120
GAM	0.338
GAMM	0.220

Table 4.2: Residuals standard deviation for all models

several points lie outside the prediction interval.

In addition to the forecasting accuracy, we evaluate the models by checking the models' residual diagnostics. The residuals from the TBATS fitted model are shown in FigureC-4. From the residuals time plot, the residuals variations remain constant apart from some outliers. However, the TBATS model is able to handle linear and non-linear time series by using Box-Cox transformation. Also, residuals are normally distributed with mean zero and 0.112 standard deviation as provided in Table 4.2. The ACF plot of the residuals illustrates that most of the lags autocorrelation fall outside the 95% confidence bounds indicating that a lot of information has not been captured with the model. To confirm this result, the Jung-Box test and Box-Pierce test are done to check whether the residuals are white noise. The results of the tests are significant since the p-value is less than 0.05. Therefore, the residuals are not white noise and there is a lot of information that has not been captured.

4.1.2 DR Model

The DR model comes after the TBATS model for the month-ahead forecast with RMSE 8.7 and ME 6.76. Compared to other models, it has lower precision for shorter forecasting horizon like week-ahead and day-ahead forecasting. However, the difference in RMSE and MAE between the DR model and other models is not high as seen in Figure4-2b and 4-2c. These results demonstrate that as the forecasting horizon becomes longer, the superiority of the DR model performance increases significantly. It is worth mentioning that our results for the DR model are compatible with those presented in [YPT99].

The prediction interval is produced to show the uncertainty associated with point forecasts. The shaded region in Figure 4-3b illustrates the prediction interval for week-ahead forecasting with the DR model. As presented in the figure, the prediction interval of the DR model is more narrow than TBATS model prediction intervals.

On the other hand, we evaluate the performance of the DR model by checking residual diagnostics. In FigureC-5, the residuals of the DR model are obtained and depicted. We find that the variance of the residuals is constant with the exception of some outliers. Moreover, the histogram plot suggests that the residuals are normally distributed with mean zero and 0.12 standard deviation as shown in Table 4.2. The ACF plot shows many significant autocorrelations lie outside the 95% confidence bounds. Also, the results of the Jung-Box test and Box-Pierce test demonstrate that there is information left in the residuals and should be used by the DR model to compute forecasts.

4.1.3 STLM Model

For a long forecasting horizon like month-ahead forecasting, the STLM model comes after the TBATS model. The difference between the STLM model and the TBATS model in RMSE and MAE is 1.39 and 1.04 respectively. Although, the STLM model outperforms the TBATS model for week-ahead forecasting as well as day-ahead forecasting, but still has higher RMSE and MAE than the GAMM model. As mentioned before, for long time series when the seasonality changes over time, STLM and TBATS are recommended. The rate of change of the seasonal pattern is controlled by the user. Also, the STLM model is robust to outliers so the sudden peaks of the electricity prices series will affect only the remainder component not the estimates of the trend-cycle nor the seasonal components.

The STLM model prediction intervals are calculated using the prediction intervals from the seasonally adjusted series while ignoring the uncertainty related to the seasonal component. Also, it ignores the uncertainty in the future seasonal pattern. Figure4-3c displays week-ahead forecasting with STLM model, where the shaded area represents the prediction interval. We observe that STLM model prediction interval is wider than other models.

As seen in FigureC-6, the residuals obtained from forecasting electricity prices using STLM model are plotted. We notice that residuals has constant variance apart from many outliers. From the histogram plot, residuals has normal distribution with mean zero and standard deviation 0.055 (as shown in Table 4.2). Moreover, the ACF plot has significant autocorrelation especially for the first lags. With p-value less than 0.05, the Jung-Box test and Box-Pierce test results in significant residuals that are distinguishable from a white noise series.

4.1.4 GAM and GAMM Models

In comparison to other models, GAM and GAMM models has lower accuracy for month-ahead forecasting. Their RMSE is higher than TBATS model by more than 1.4 and their MAE is higher by more than 1.4. In contrast to the month-ahead forecasting performance, GAMM and GAM models have the best performance among other models in week-ahead forecasting as provided in Table4.1b. Similarly, they outperform all models in day-ahead forecasting. Therefore, over longer forecasting horizons including month-ahead forecasting, GAMM and GAM models are found to be less accurate and reliable than other models.

In Figure C-7 and Figure C-8, residual diagnostics plots are produced for both GAM and GAMM model. The residuals are normally distributed for both models with standard deviation of the GAM model being higher than the GAMM model standard deviation as presented in Table4.2. The ACF plot for the GAM model displays significant autocorrelation with the Jung-Box test and Box-Pierce test both confirm that residuals are not white noise and there is still information has not been captured. So, we fit ARIMA model to the GAM model errors. The ACF plot of the ARIMA model errors shows small autocorrelation for few lags that lie outside the 95% confidence bounds. But the results of the Jung-Box test and Box-Pierce test show that the residuals are not distinguishable from a white noise series.

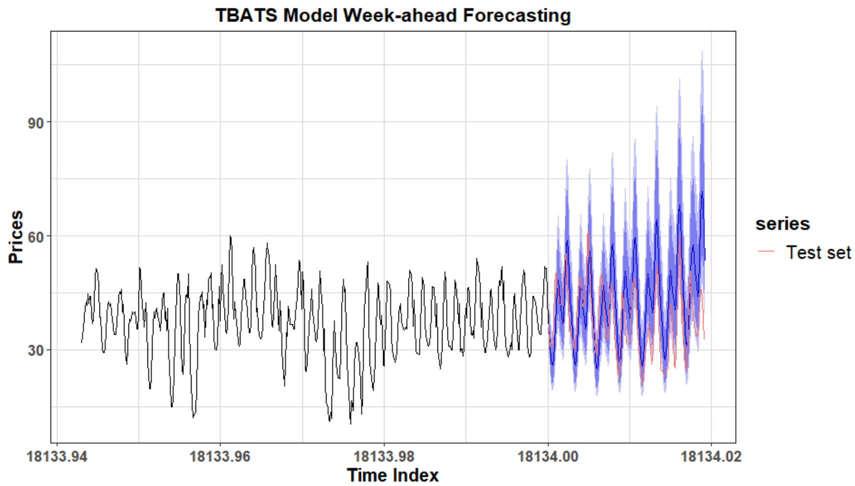
4.1.5 RF Model

The RF model is the only machine learning model used in this thesis. For month-ahead forecasting, RF model has the higher RMSE and MAE than most of the models used. But the difference in RMSE and MAE between the models is small as seen in Figure4-2. For week-ahead and day-ahead forecasting, the performance of the RF model improves compared to other models as seen in Table4.1b and Table4.1c. Our day-ahead forecasting results is comparable with those presented in [LRS18]. The RF model might be a good method for producing optimal forecasts when using more predictors like weather information as they could improve the performance.

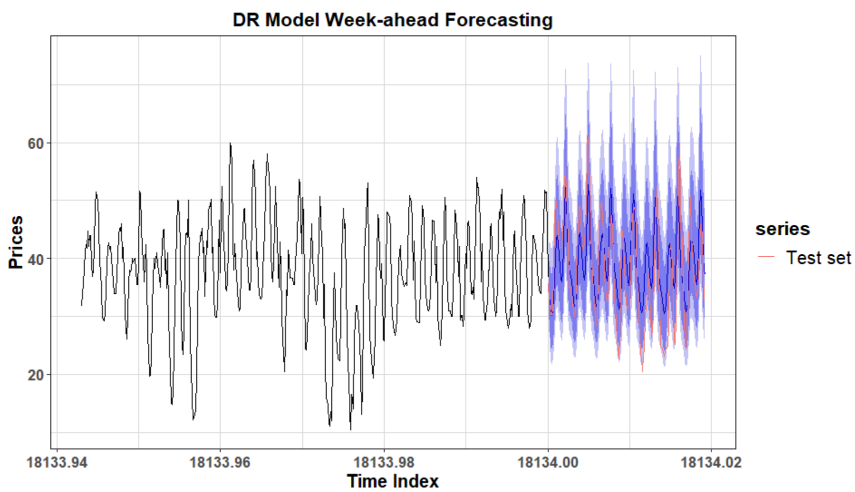
4.2 DM Test

In this section, we study the statistical significance of the difference in forecasting accuracy between all considered models. By applying the DM test steps as defined with equations (B.12),(B.13) and (B.14) in section B.2, we obtain the results which are summarized in Table 4.3a, Table 4.3b and Table 4.3c. The $\surd$ symbol means the forecasting accuracy of M_1 is statistically significantly better than M_2 forecasting accuracy when considering the full loss differential. The empty cells mean M_2 is at least significantly better than M_1 in one hour of the forecasting horizon (day or week or month). Based on the results represented in Table 4.3a, 4.3b and 4.3c, several observations can be confirmed:

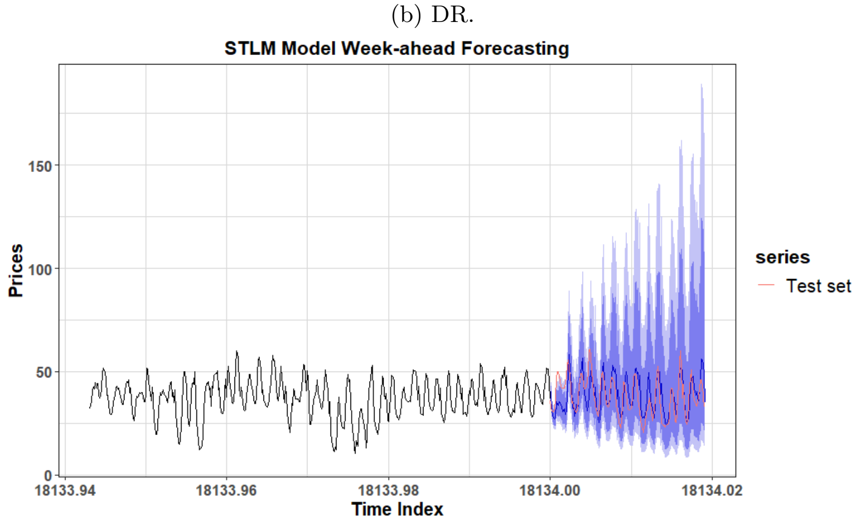
- For month-ahead forecasting, the TBATS model forecasting accuracy is statistically significantly better than other models.
- The DR model has significantly better forecasting accuracy than all models except for the TBATS model for month-ahead forecasting.
- For week-ahead forecasting, TBATS model and GAM model forecasting accuracy are significantly better than other models .The accuracy of GAM and TBATS model are not significantly different.
- For day-ahead forecasting, TBATS and GAM models forecasting accuracy is better than other model.
- We can observe that all models are significantly better than the RF model for all forecasting horizons.



(a) TBATS.



(b) DR.



(c) STLM.

Figure 4-3: Example of Weekly forecast with prediction intervals

M_1 \ M_2	**STLM**	**TBATS**	**DR**	**GAM**	**RF**
STLM					√
TBATS	√		√		√
DR					√
GAM	√		√		√
RF					

(a) Day-ahead.

M_1 \ M_2	**STLM**	**TBATS**	**DR**	**GAM**	**RF**
STLM					√
TBATS	√		√		√
DR	√				√
GAM	√		√		√
RF					

(b) Week-ahead.

M_1 \ M_2	**STLM**	**TBATS**	**DR**	**GAM**	**RF**
STLM				√	√
TBATS	√		√	√	√
DR	√			√	√
GAM					√
RF					

(c) Month-ahead.

Table 4.3: Diebold-Mariano results

4.3 Discussion

Forecasting electricity spot prices in the wholesale electricity markets is one of the important decisions to be made in everyday operations and long-term planning. The goal of this thesis is to forecast electricity spot prices with different statistical and machine learning methods for different forecasting horizons. Overall, we observed that time series cross validations make more efficient use of the available data and provide more robust estimations than out-of-sample method. In addition, for all considered models we noticed that the shorter the forecasting horizons, the more precise the forecast tends to be. Also, longer forecasting horizons lead to increasing errors and more uncertainties are involved than short forecasting horizons. However, the higher forecasting error, should not be attributed to poorer model performance but rather to the longer period over which it was evaluated.

Our results demonstrate that the improved shorter horizon forecasting accuracy does not always indicate better accuracy for longer forecasting horizons. Furthermore, the produced accuracy in terms of average RMSE in the range 4-28 for month-ahead forecasting, 4-13 for week-ahead forecasting and 2-13 for day-ahead forecasting. Thus, more research has to be done to improve the accuracy. Most of the forecasting models don't work well for long time series such as the electricity prices series. As the number of observations increase, parameters optimization becomes computationally expensive. This problem could be solved by using non-parametric models (ML models) that would work with the long time series. But one drawback of using machine learning models for time series forecasting is that we can't measure the predictions' uncertainty in terms of frequentist confidence or Bayesian credible interval. Also, failing the Jung-Box and Box-Pierce tests don't necessarily mean that the model results in poor forecasts nor that the prediction intervals are imprecise. In fact, it implies that there is more information in the electricity prices data than is captured in the model and might not affect model performance much. Finally, as with all modelling to get the most accurate results, extensive forecasting is desirable however this is computationally costly and takes a long time to complete.

Chapter 5

Conclusion

In this chapter, we summarize the results of our analysis and outline possible directions for future work.

5.1 Summary

In this dissertation, we studied forecasting techniques for electricity prices. We focused on analyzing the performance of various techniques for different forecasting horizons. The analyses were performed in R.

We started by studying several time series methods including TBATS, STLM, DR, GAM, GAMM and RF model. We evaluated the performance of these methods for different forecasting horizons such as day-ahead, week-ahead and month-ahead. We used historical electricity prices as the main predictor. Also, calendar effects were utilized to forecast electricity prices such as hour of the day, day of the week, day of the year, month of the year and whether the day a normal weekday, a weekend or a public holiday. The performance of these methods was examined using time series cross-validation in terms of root mean square (RMSE) and mean absolute error (MAE). Our results showed that GAMM model has the best performance in day-ahead forecasting as well as week-ahead forecasting in terms of RMSE and MAE. Also, as the forecasting horizons were increased the forecasting error also increased. However, we found that the TBATS model is more robust against increasing forecasting horizons than other models. For month-ahead forecasting, the TBATS model has the best forecasting accuracy compared to other models. The TBATS model proved to have statistically significantly better accuracy than all considered models. While in both, week-ahead and day-ahead forecasting, TBATS and GAM models have statistically significantly better forecasting accuracy compared to other models. Most of the models are not significantly different in day-ahead forecasting.

In summary, our results are consistent and have improved on the previous investigations into the impact of forecasting horizons on the performance of different models used for forecasting electricity prices. In particular, unlike other investigations, we thoroughly investigated three forecasting horizons instead of focusing on day-ahead forecasting. Furthermore, no other study looked at the impact of week-ahead and month-ahead forecasting on different models' performance in forecasting electricity prices for the electricity wholesale market in the UK.

5.2 Future Work

There is a variety of interesting topics for future research. Here are few of them:

Incorporating the other influential factors: Factors on electricity prices include weather conditions,market's design and transmission congestion. It would also be interesting to investigate the impact of including these factors on models' performance for different forecasting horizons.

Hybrid models: A natural extension to our work is to use a combination of models to forecast electricity prices for different forecasting horizons and examine if the performance would improve.

Machine learning models: Various machine learning models can be used to forecast electricity prices such as an extreme gradient boosting (XGB) model. Investigate the performance of many machine learning models in electricity prices forecasting while including more influential factors.

Time series cross validation: It would be interesting to use different time series cross validations methods and compare between them to see if any of them would enhance our models performance.

Long-term forecasting horizon: A natural extension to our work also is examining the performance of our models in long-term electricity prices forecasting.

Deep learning Models: It would be useful to use deep learning models for predicting electricity prices and investigate if they lead to significant improvements in the forecasting accuracy.

Appendix A

Forecasting Models and Equations

A.1 Seasonal ARIMA Model

The ARMA(p,q) model is based on the idea of the current price value x_t can be explained linearly as a function of p past values, x_{t-1}, x_{t-2},x_{t-3},...,x_{t-p} (autoregressive part), and in terms of the linearly combined q past forecast error (moving average part). It can be represented as:

$$\phi(B)X_t = \theta(B)w_t \tag{A.1}$$

where:

B : backward shift operator i.e.,$B^h X_t \equiv X_{t-h}$.

$\phi(B)$: shorthand for $\phi(B)= 1 - \phi_1 B - \cdots - \phi_p B^p$ where $\phi_1, \cdots, \phi_p$ are coefficients of autoregressive

$\theta(B)$: shorthand for $\theta(B)=1 + \theta_1 B + \cdots + \theta_q B^q$ where $\theta_1, \cdots, \theta_q$ are coefficients of moving average polynomials.

w_t : white noise with zero mean and finite variance №$(0, \sigma^2)$.

When (p=0) we get the moving average model MA(q) and when (q=0) we get the autoregressive model AR(p). In addition to ARMA model parameters, ARIMA model has the number of differencing involved at lag-one (d) and can be denoted as ARIMA(p,d,q). ARIMA model can be written as [ZYN+06]:

$$\phi(B)\nabla^d X_t = \theta(B)w_t \tag{A.2}$$

where:

$\nabla x_t \equiv (1 - B)x_t$: lag-1 differencing operator.

$\nabla_h x_t \equiv (1 - B^h)x_t \equiv x_t - x_{t-h}$: lag-h differencing operator.

Suppose the electricity prices series $y(t)$ is a non-stationary stochastic process as $y(t) = y_1, y_2, ..., y_n$, which can be described with:

$$y(t) = S(t) + T(t) + R(t) \tag{A.3}$$

where:

$S(t)$: a periodical component known as seasonal component.

$T(t)$: non-periodic trend component .

$R(t)$: a random noise component reflecting random fluctuation also know as the remainder component and it is usually represented as a stationary time series.

From the above equation, the remainder component can be obtained by eliminating the trend and the seasonal components with the following process:

- The trend is eliminated by applying differencing. If there is no seasonal component the differencing can be repeated until a stationary process is obtained. The dth order difference operator $\nabla^d y_t = \nabla(\nabla^{d-1})y_t$ targets to extract dth-order trend component.

- The seasonal component is eliminated by applying seasonal differencing. If the hourly electricity price time series has one periodic component with one day as a period by applying $(1 - B^2 4)y_t$ the electricity price series can be transformed to a stationary series.

The above process can be implemented for the electricity price series $y(t)$ by ARIMA model. After applying the above two operations the remainder component $R(t)$ is obtained. When the time series is a complex non-stationary process, ARIMA(P,D,Q) can be represented as:

$$\phi(B^s)\nabla_s^D y_t = \theta(B^s)Z_t \tag{A.4}$$

where:

s: the time series y_t period.

D: the order of the trend.

$\nabla_s = (1 - B^s)$.

${\nabla_s}^D = (1 - B^s)^D$.

$\phi(B^s) = 1 - \phi_1 B^s - \phi_2 B^2 s - \cdots - \phi_p B^P s$.

$\theta(B^s) = 1 - \theta_1 B^s - \theta_2 B^2 s - \cdots - \theta_Q B^Q s$.

$\phi_1, \phi_2, ..., \phi_P and \theta_1, \theta_2, ..., \theta_Q$: constants.

P, Q: orders of $\phi(B^s)$ and $\theta(B^s)$.

Z_t is another ARIMA(p,d,q) model can be represented as before:

$$\phi(B)\nabla^d Z_t = \theta(B)R_t \tag{A.5}$$

By joining the last two equations, the result can be written as :

$$\phi(B)\phi(B^s)\nabla^d\nabla^{sD}y_t = \theta(B)\theta(B^s)R_t \tag{A.6}$$

This equation represents the seasonal ARIMA model with the order (p,d,q)X(P,D,Q).

A.2 TBATS Model

If the time series at t is y_t, $y_t^{(\lambda)}$ is the time series at time t after Box-Cox transformation. After time series transformation, decomposition results into level component l_t, growth component b_t, an irregular component d_t, and seasonal component $S_t^{(i)}$ with seasonal frequencies m_i. TBATS model components are described as follows:

$$y_t^{\lambda} = l_{t-1} + \phi b_{t-1} + \sum_{i=1}^{T} S_{t-m_i}^{(i)} + d_t \tag{A.7}$$

$$l_t = l_{t-1} + \phi b_{t-1} + \alpha d_t \tag{A.8}$$

$$b_t = \phi b_{t-1} + \beta d_t \tag{A.9}$$

$$d_t = \sum_{i=1}^{P} \varphi_i d_{t-i} + \sum_{i=1}^{q} \theta_i R_{t-i} + R_t \tag{A.10}$$

Where:

$y_t^{(\lambda)}$: Time series at time t (after box-cox transformation).

$S_t^{(i)}$: ith seasonal component.

l_t : Local level component.

b_t : Trend with damping or growth component.

d_t : ARMA(p,q) process for residual.

R_t : Gaussian white noise with zero mean and constant variance σ.

T : Amount of seasonality.

m_i : Length of ith seasonal period.

k_i : Number of harmonics for ith seasonal period.

λ : Box-Cox transformation parameter.

α, β : Smoothing parameters.

ϕ : Trend damping parameter.

φ_i, θ_i : ARMA(p,q) coefficients.

$\gamma_1^{(i)}, \gamma_2^{(i)}$: Seasonal smoothing (two for each period.

The seasonal part can written as:

$$S_t^{(i)} = \sum_{j=1}^{(k_i)} S_{j,t}^{(i)} \tag{A.11}$$

$$S_{j,t}^{(i)} = S_{j,t-1}^{(i)} cos(\omega_i) + S_{j,t-1}^{*(i)} sin(\omega_i) + \gamma_1^{(i)} d_t \tag{A.12}$$

$$S_{j,t}^{*(i)} = -S_{j,t-1}^{(i)} sin(\omega_i) + S_{j,t-1}^{*(i)} cos(\omega_i) + \gamma_2^{(i)} d_t \tag{A.13}$$

$$\omega_i = 2\pi_j / m_i \tag{A.14}$$

A.3 ETS Model

The ETS Model [HA18][Cog18] is Error, Trend and Seasonal which is an exponential smoothing state space model that is composed of a measurement equation to represent the observed data and state equations to represent the unobserved states (level,trend,seasonal) that change over time. Different exponential smoothing methods are produced by changing the combinations of the trend component, seasonal component and errors whether it is additive or multiplicative errors. Each state model is labeled as ETS (Error,Trend,Seasonal). The model used in this thesis is ETS(A,N,N) where:

A: additive errors.

N: None Trend.

N: None Seasonal.

If there is a time series data up to time $t-1$, and the next observation y_t is forecasted as $\hat{y}_t$, the forecast error:

$$e_t = y_t - \hat{y}_t \tag{A.15}$$

To forecast the time series at $t+1$, the simple exponential smoothing takes the forecast of the previous period and adjusts with the forecast error as:

$$\hat{y}_{t+1} = \hat{y}_t + \alpha(y_t - \hat{y}_t) \tag{A.16}$$

where α is constant between 0 and 1. The above equation can be written as:

$$\hat{y}_{t+1} = \alpha y_t + (1-\alpha)\hat{y}_t \tag{A.17}$$

$\hat{y}_{t+1}$ can be represented as a weighted moving average of all previous observations with the weights decreasing exponentially as follows:

$$\hat{y}_{t+1} = \alpha y_t + \alpha(1-\alpha)y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \cdots + \alpha(1-\alpha)^{t-1} y_1 + \alpha(1-\alpha)^t \hat{y}_1 \tag{A.18}$$

The choice of the first point is very important and known as initialization problem ($\ell_0 = \hat{y}_1$). Also, the forecast function is assumed to be flat so:

$$\hat{y}_{t+h|t} = \hat{y}_{t+1}, h = 2, 3, ... \tag{A.19}$$

$$\ell_t = \hat{y}_{t+1} \quad \text{Forecast equation} \tag{A.20}$$

$$\ell_t = \alpha y_t + (1-\alpha)\ell_{t-1} \quad \text{Smoothing equation} \tag{A.21}$$

$$\hat{y}_{t+h|t} = \ell_t \tag{A.22}$$

ℓ_t is a measure of the level of the series at time t. The above equations can also be written as:

$$\ell_t = \ell_{t-1} + \alpha(y_t - \ell_{t-1}) \tag{A.23}$$

$$\ell_t = \ell_{t-1} + \alpha e_t \tag{A.24}$$

where e_t is the residual at time t which is called error correction. Models with additive errors assume the residuals are normally distributed white noise with mean 0 and variance σ^2.

$$e_t = \epsilon_t \sim \mathcal{N}(\mu, \sigma^2) \tag{A.25}$$

A.4 GAM and GAMM Model

Smooth functions can be determined using smooth splines which are piecewise polynomial functions (basis functions) that connect many polynomials to make a smooth curve. The set of points that connect the polynomial pieces are called knots. The connected polynomials at each knots share the same derivatives up to a given degrees. The level of model smoothness is controlled by the number, location of knots, the degree of the polynomials as well as the penlty parameter λ. The smooth function can be determined by minimizing the penalized sum of squares:

$$\underbrace{\sum_{i=1}^{n}[y_i - f(x_i)]^2}_{\text{first part}} + \underbrace{\lambda \int_{x_{min}}^{x_{max}} [f''(x)]^2 dx}_{\text{Second part}} \tag{A.26}$$

Where:

First part : the standard residual sum of squares, which describes how the fitted values are close to the observed values.

Second part : penalizes models that are too unsmooth.

λ: smoothing parameter controls the trade off between model fit and model smoothness.

In [Woo17], the best option for the polynomial degree of the smoothing term is the cubic smoothing splines where knots are located at every data point. When the number of knots are equal to the data points, if λ is not chosen appropriately then model overfitting may result. Also, it will result in to high computational cost. Another option to controlling smoothness rather than changing the basis dimension, is by adding "wiggliness" penalty to the least square fitting objective. In this case the smooth function is represented as penalised regression spline [Woo17]. It can be expressed as:

$$s(x_j) = B_0(x_j)\beta_0 + B_1(x_j)\beta_1 + \cdots + B_q(x_j)\beta_q = \boldsymbol{B'\beta}, \forall j = 1, \cdots, p \tag{A.27}$$

Where:

$B_0(.), \cdots, B_q(.)$: basis functions.

$\beta_0, \cdots, \beta_q$: associated coefficients with the basis dimension q and act as an amplifier of the curvature of the spline.

$\boldsymbol{B}$: Model matrix of the basis functions.

$\boldsymbol{\beta}$: vector of regression coefficients .

Instead of fitting the model by minimizing the least square, it could be fit by minimizing:

$$\underbrace{\|y - \beta\boldsymbol{X}\|^2}_{\text{Least Square part}} + \lambda \underbrace{\int_0^1 [f''(x)]^2 dx}_{\text{Penalty Part}} \tag{A.28}$$

Where: λ: smoothing parameter that control the trade off between model fit and model smoothness.

Penalty part: integrated square of second derivative penalizes model.

The penalty part can be written as quadratic form in β:

$$\int_0^1 [f''(x)]^2 dx = \beta^T \boldsymbol{P} \beta \tag{A.29}$$

Where $\boldsymbol{P}$ is is a matrix of known coefficients.Thus, the penalized least squares estimator of β is:

$$\hat{\beta} = (\boldsymbol{X^T X} + \lambda \boldsymbol{P})^{-1} \boldsymbol{X^T} y \tag{A.30}$$

The smoothness is imposed by directly penalizing the difference among the adjacent coefficients. This method is called the Penalised Iterativley Reweighted Least Square method (P-IRLS). By using P-IRLS , the regression coefficients $\hat{\beta}$ can be calculated for any λ. Very large value of $\lambda \to \infty$ results in a straight line estimate for f, while small values of $\lambda \to 0$ lead to an un-penalized spline estimate. One way of determining the optimal value of λ(degree of smoothness) is the backwards selection method. This method can result in poor model accuracy if the locations of the knots are unevenly-spaced. Also it's computationally expensive. An alternative methods for λ estimation are the mixed model approach via Restricted Maximum Likelihood (REML) or the Generalized Cross Validation criteria (GCV).

Appendix B

Model Evaluation

B.1 Accuracy measures

The time series data $y_1, \ldots, y_T$ is split to training data $y_1, \ldots, y_N$ to estimate models and testing data $y_{N+1}, \ldots, y_T$ to evaluate model performance.

$$e_t = y_t - \hat{y}_t, t = N + 1, ...T \tag{B.1}$$

$$p_t = \frac{100e_t}{y_t} \tag{B.2}$$

where:

e_t:The forecast error.

p_t :The percentage error.

$\hat{y}_t$: the forecasted value.

$$sMAPE = \frac{2|e_i|}{|y_t| + |\hat{y}_t|} \tag{B.3}$$

When the y_t is close to zero the forecasted value $\hat{y}_t$ is also close to zero. The division by small values makes sMAPE unstable.

$$ME = mean(e_i) \tag{B.4}$$

$$RMSE = \sqrt{mean(e_i^2)} \tag{B.5}$$

$$MAE = mean(|e_i|) \tag{B.6}$$

$$MPE = mean(p_t) \tag{B.7}$$

$$MAPE = mean(|p_t|) \tag{B.8}$$

$$q_t = e_t/Q \tag{B.9}$$

where Q is a scaling statistic computed on the training data. For non-seasonal time series Q can be defined as the MAE for naïve forecasts calculated on the training data. For seasonal time series with period m, Q can be defined as:

$$q_i = \frac{e_i}{\frac{1}{T-m} \sum_{t=m+1}^{T} |y_t - y_{t-m}|} \tag{B.10}$$

$$MASE = mean(|q_i|) = \frac{MAE}{Q} \tag{B.11}$$

B.2 Diebold-Mariano test

We applied the DM test for each hour and for model pairs M_1,M_2 following the steps in cite[2]:

- At 95% confidence level ,we perform one sided test with null hypothesis of M_1 forecasting accuracy is equal or less than M_2 forecasting accuracy:

$$DM_h \begin{cases} H_0 : \mathbb{E}[d_{h,k}^{M_1,M_2}] \geq 0, \\ H_1 : \mathbb{E}[d_{h,k}^{M_1,M_2}] < 0, \end{cases} \quad \text{for h=1,2,....} \tag{B.12}$$

  Where: $d_k^{M_1,M_2} = |\varepsilon_k^{M_1}| - |\varepsilon_k^{M_2}|$ and $[d_{h,1}^{M_1,M_2}, ..., d_{h,N/24}^{M_1,M_2}]$

- We perform the one sided DM test with null hypothesis of M_2 forecasting accuracy is equal or less than M_1 forecasting accuracy:

$$\hat{DM}_h \begin{cases} H_0 : \mathbb{E}[-d_{h,k}^{M_1,M_2}] \geq 0, \\ H_1 : \mathbb{E}[-d_{h,k}^{M_1,M_2}] < 0, \end{cases} \quad \text{for h=1,2,....} \tag{B.13}$$

  The forecasting accuracy of M_1 is better than M_2,if at least one of the DM_h null hypothesis is rejected and non of the $\hat{DM}_h$ null hypothesis is rejected.

- If at least one of the DM_h and $\hat{DM}_h$ null hypothesis is rejected, we carry out a DM test on the full loss differential $[d_1, ..., d_N]$ taking into account the serial correlation.

$$DM_{sc} \begin{cases} H_0 : \mathbb{E}[d^{M_1,M_2}] \geq 0, \\ H_1 : \mathbb{E}[d^{M_1,M_2}] < 0, \end{cases} \tag{B.14}$$

  Rejecting the DM_{sc} null hypothesis indicated that, M_1 forecasting accuracy is significantly better than M_2 although at some hours M_2 forecasting accuracy is significantly better than M_1.

Appendix C

Figures

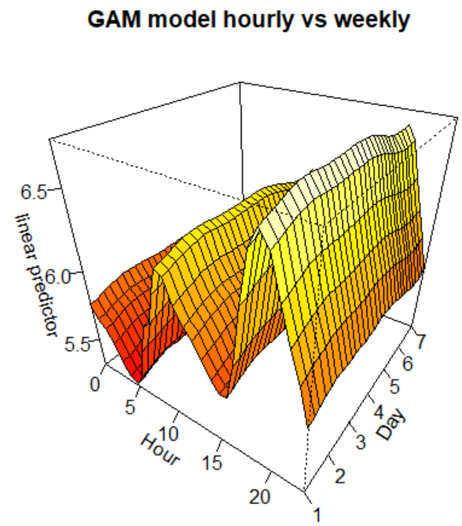


Figure C-1: GAM model hour vs week.

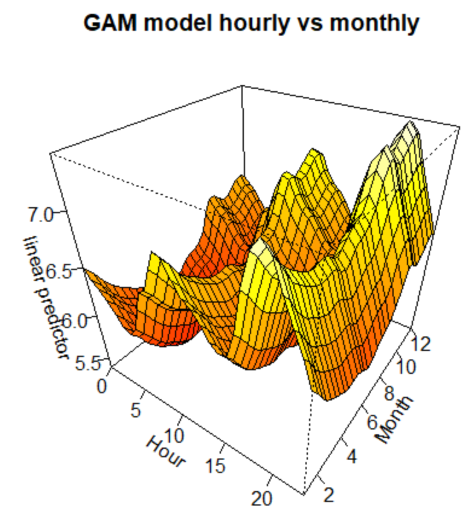


Figure C-2: GAM model hour vs month.

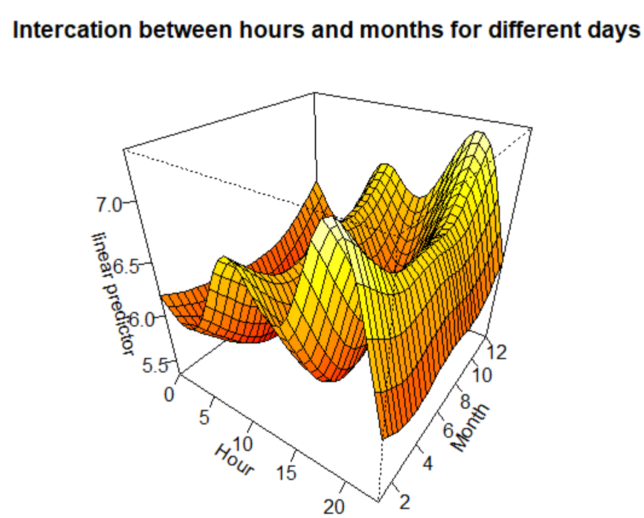


Figure C-3: GAM the interaction term (hour,week,month)

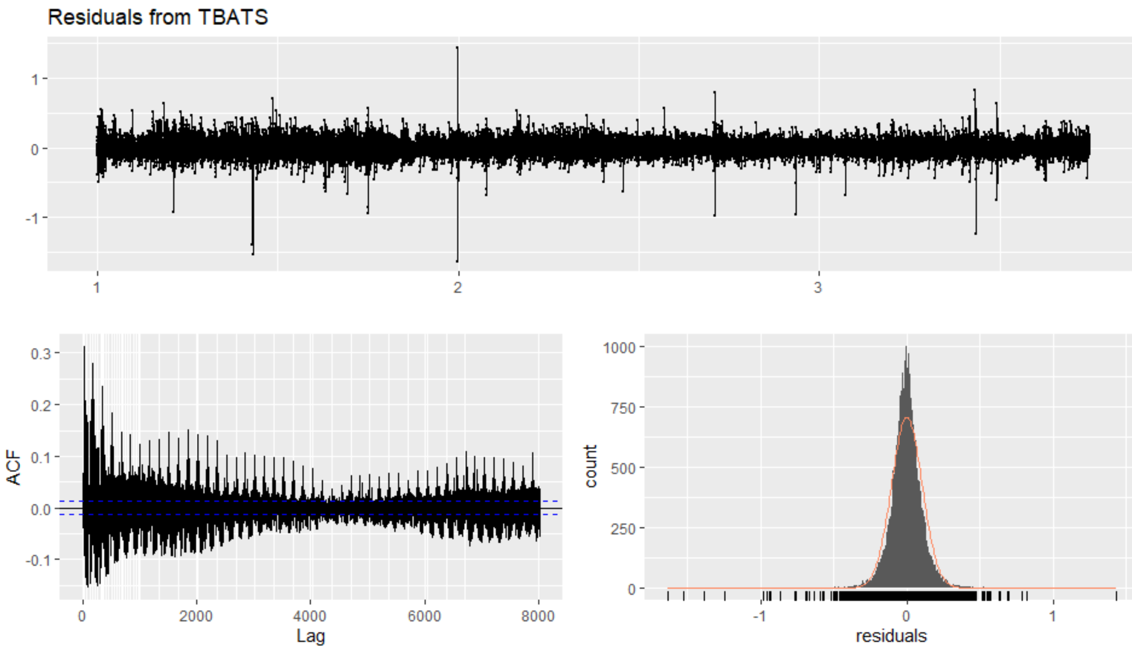


Figure C-4: Residuals from TBATS Model.

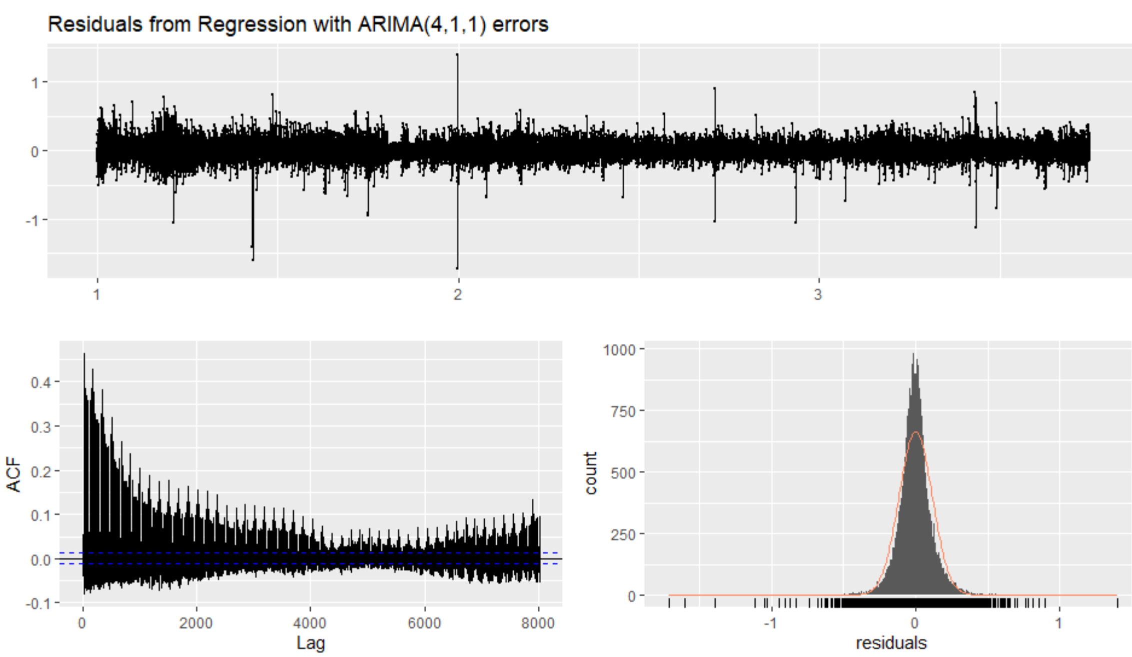


Figure C-5: Residuals from DR Model.

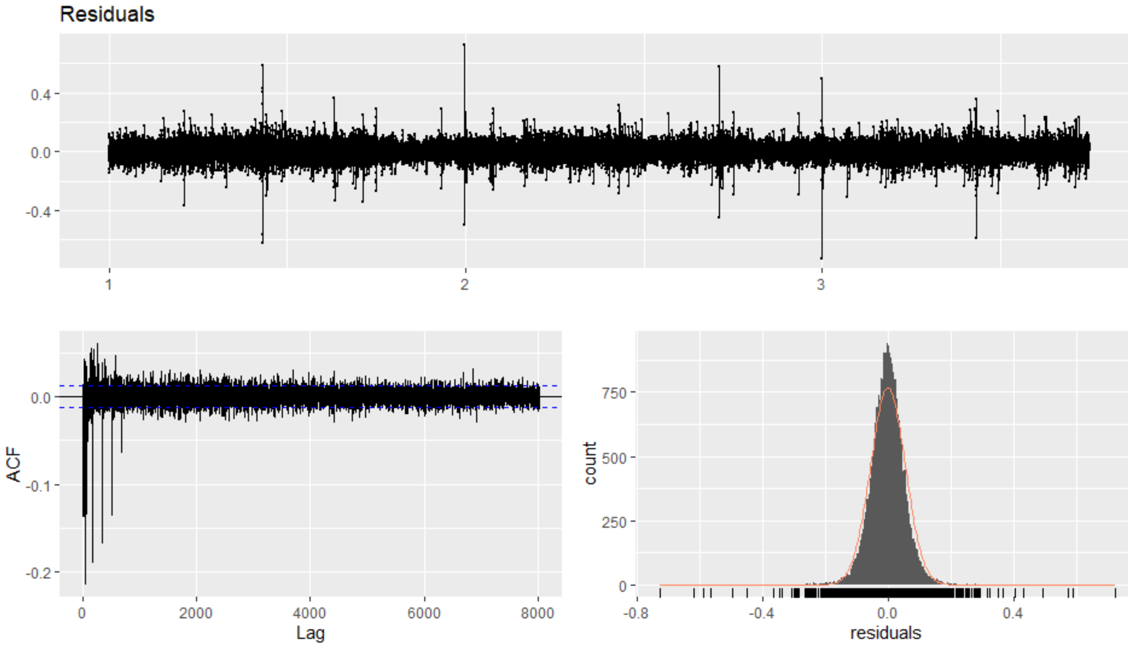


Figure C-6: Residuals from STLM Model.

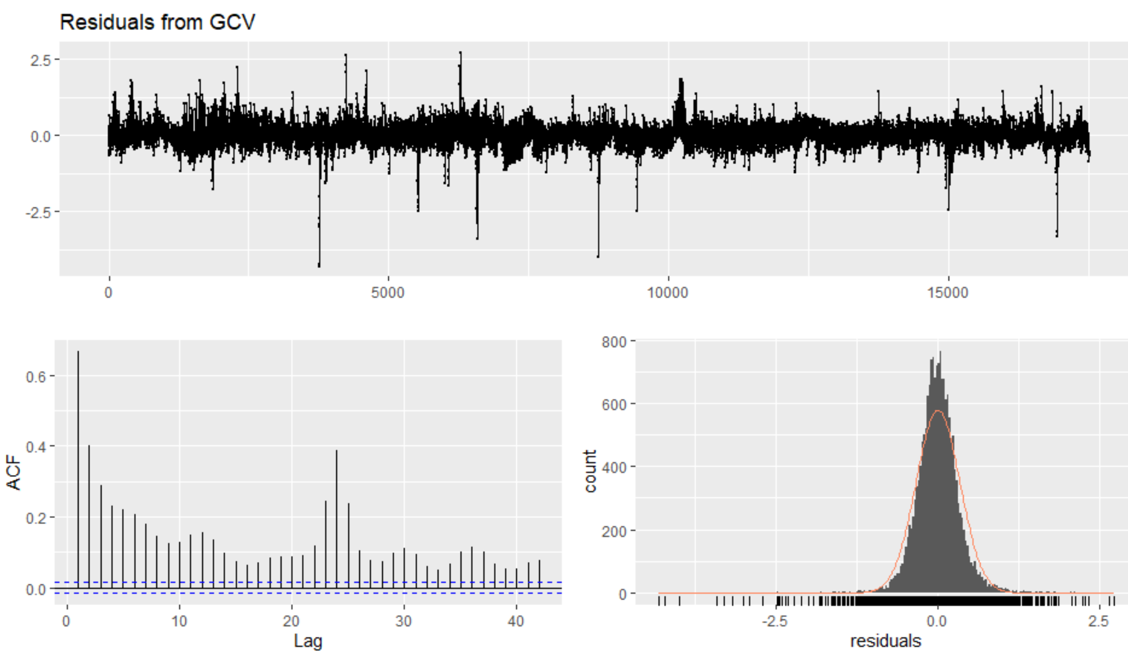


Figure C-7: Residuals from GAM Model.

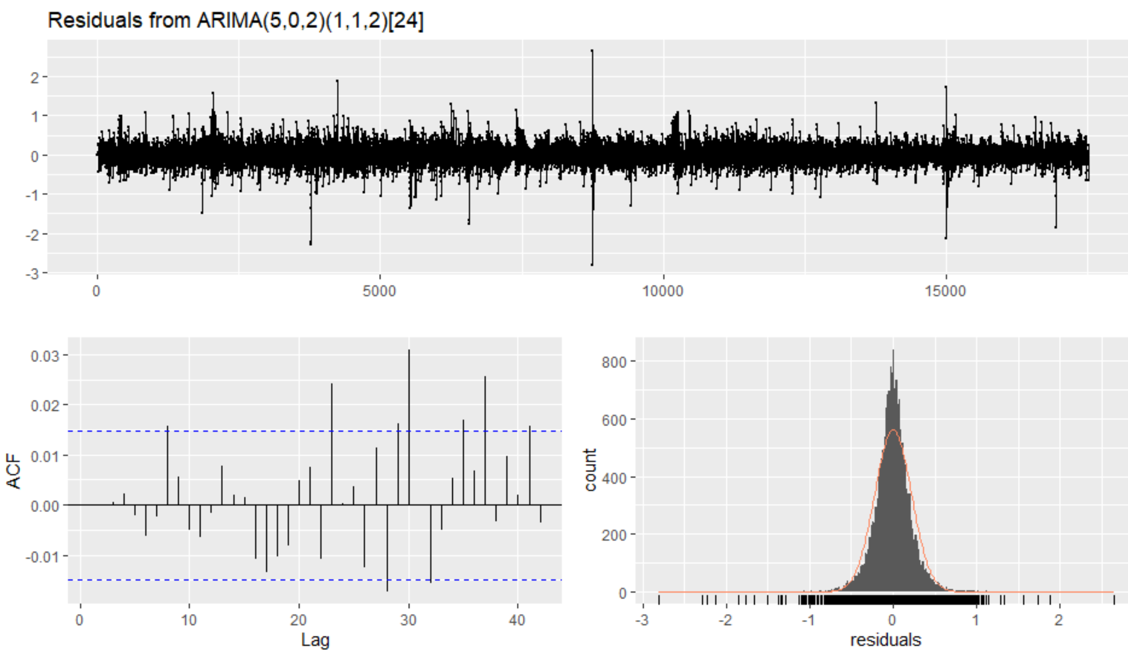


Figure C-8: Residuals from GAMM Model.

Bibliography

[ASK09] Sahil Kumar Aggarwal, Lalit Mohan Saini, and Ashwani Kumar. Electricity price forecasting in deregulated markets: A review and evaluation. 2009.

[BJRL15a] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[BJRL15b] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[CCMT90] Rb Cleveland, William S. Cleveland, Jean E. McRae, and Irma J. Terpenning. Stl: A seasonal-trend decomposition procedure based on loess (with discussion). 1990.

[Cog18] Avril Coghlan. A little book of r for time series. 2018.

[DM02] Francis X Diebold and Robert S Mariano. Comparing predictive accuracy. *Journal of Business & economic statistics*, 20(1):134–144, 2002.

[DW52] Field Nathan David and Peter Whittle. Hypothesis-testing in time series analysis. 1952.

[HA18] Robin John Hyndman and George Athanasopoulos. *Forecasting: Principles and Practice*. OTexts, Australia, 2nd edition, 2018.

[HK06] Rob J. Hyndman and Anne B. Koehler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679 – 688, 2006.

[HTF09] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics)*. 02 2009.

[HTWI09] L. Hu, G. Taylor, H. Wan, and M. Irving. A review of short-term electricity price forecasting techniques in deregulated electricity markets. In *2009 44th International Universities Power Engineering Conference (UPEC)*, pages 1–5, 2009.

[Hyn14] Rob J Hyndman. Measuring forecast accuracy. *Business forecasting: Practical problems and solutions*, pages 177–183, 2014.

[LHS10] Alysha Livera, Rob Hyndman, and Ralph Snyder. Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, 106:1513–1527, 01 2010.

[LRS18] Jesus Lago, Fjo [De Ridder], and Bart [De Schutter]. Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. *Applied Energy*, 221:386 – 405, 2018.

[LRVS18a] Jesus Lago, Fjo [De Ridder], Peter Vrancx, and Bart [De Schutter]. Forecasting day-ahead electricity prices in europe: The importance of considering market integration. *Applied Energy*, 211:890 – 903, 2018.

[LRVS18b] Jesus Lago, Fjo [De Ridder], Peter Vrancx, and Bart [De Schutter]. Forecasting day-ahead electricity prices in europe: The importance of considering market integration. *Applied Energy*, 211:890 – 903, 2018.

[Mak93] Spyros Makridakis. Accuracy measures: theoretical and practical concerns. *International Journal of Forecasting*, 9(4):527–529, 1993.

[MBB17] Steffen Moritz and Thomas Bartz-Beielstein. imputeTS: Time Series Missing Value Imputation in R. *The R Journal*, 9(1):207–218, 2017.

[MSL19a] Jan-Hendrik Meier, Stephan Schneider, and Chan Le. Short-term electricity price forecasting using generalized additive models. In *ICTERI Workshops*, 2019.

[MSL19b] Jan-Hendrik Meier, Stephan Schneider, and Chan Le. Short-term electricity price forecasting using generalized additive models. In *ICTERI Workshops*, 2019.

[NMI18] Iram Naim, Tripti Mahara, and Ashraf Rahman Idrisi. Effective short-term forecasting for daily time series with complex seasonal patterns. *Procedia Computer Science*, 132:1832 – 1841, 2018. International Conference on Computational Intelligence and Data Science.

[STMY15] Nitin Singh, Shailesh Tiwari, S. R. Mohanty, and Deepak Yadav. A modified cost function generalized neuron for electricity price forecasting in deregulated power markets. In *Proceedings of the Sixth International Conference on Computer and Communication Technology 2015*, ICCCT '15, page 120–126, New York, NY, USA, 2015. Association for Computing Machinery.

[VGT+14] Andreas Veit, Christoph Goebel, Rohit Tidke, Christoph Doblander, and Hans-Arno Jacobsen. Household electricity demand forecasting – benchmarking state-of-the-art methods. 04 2014.

[Wer14] Rafał Weron. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4):1030 – 1081, 2014.

[Woo17] Simon N Wood. *Generalized additive models: an introduction with R*. CRC press, 2017.

[YPT99] Peter C. Young, Diego J. Pedregal, and Wlodek Tych. Dynamic harmonic regression. *Journal of Forecasting*, 18(6):369–394, 1999.

[ZYN+06] M. Zhou, Z. Yan, Y. X. Ni, G. Li, and Y. Nie. Electricity price forecasting with confidence-interval estimation through an extended arima approach. *IEE Proceedings - Generation, Transmission and Distribution*, 153(2):187–195, 2006.