CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Bachelor thesis

2017 **Štěpán Svoboda**

CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Štěpán Svoboda

Short-term electricity price forecasting - evaluation of selected hybrid models

Bachelor thesis

Prague 2017

Author: Štěpán Svoboda

Supervisor: doc. PhDr. Ladislav Krištoufek, Ph.D.

Academic Year: 2016/2017

Bibliographic note

SVOBODA, Štěpán. *Short-term electricity price forecasting - evaluation of selected hybrid models* Prague 2017. 51 pp. Bachelor thesis (Bc.) Charles University, Faculty of Social Sciences, Institute of Economic Studies. Thesis supervisor doc. PhDr. Ladislav Krištoufek, Ph.D.

Abstract

In this thesis a thorough study of the previous literature and the division and special aspects of EPF was carried out. Then the evaluation and comparison of several models was done – the ARIMA, SVR, SVRARIMA and PSF model. This comparison was done on the intra-day Nord Pool market, which is quite unique as almost all short-term EPF is carried out on the day-ahead market. Our results are robust as the modeling was done on 100 test periods and we have tested the difference in predictive accuracy using the modified DM test. Our conclusion is the PSF model is inadequate in our intra-day setting and the overall ARIMA model seems to outperform the SVR and SVRARIMA model somewhat. The dominance of ARIMA is not very strong and a further investigation of the causes of these results can better illuminate the strengths and weaknesses of these models.

Keywords

Electricity price forecasting, Support Vector Regression, Autoregressive Integrated Moving Average, SVRARIMA, Pattern Sequence-based Forecasting, Nord Pool

Abstrakt

V této práci je zevrubně popsána existující literatura na téma předpovídání ceny elektřiny, její dělení a nestandardní aspekty. Zároveň byly srovnány některé modely – ARIMA, SVR, SVRARIMA a PSF. Toto srovnání jsme provedli na vnitrodenních datech trhu Nord Pool. Tento přístup je výjimečný, protože téměř ke každému krátkodobému modelování cen elektřiny jsou využita data z trhu na jeden den dopředu. Naše výsledky jsou robustní, protože jsme použili 100 testovacích období a rozdíly v přesnosti předpovídání jsme otestovali modifikovaným DM testem. Náš závěr je, že PSF model je nevhodný k tomuto typu předpovídání a ARIMA model je celkově nejlepší. Jak SVR tak SVRARIMA model jsou marginálně horší. Rozdíl mezi těmito modely není velký a další studie, zaměřené na identifikaci důvodů těchto rozdílů, mohou přinést lepší pochopení chování těchto modelů.

Klíčová slova

Předpovídání cen elektřiny, Support Vector Regression, Autoregressive Integrated Moving Average, SVRARIMA, Pattern Sequence-based Forecasting, Nord Pool

Declaration of Authorship

I hereby proclaim that I wrote my bachelor thesis on my own under the leadership of my supervisor and that the references include all resources and literature I have used.

I grant a permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, 17 May 2017

Signature

Acknowledgment

I would like to thank my supervisor doc. PhDr. Ladislav Krištoufek, Ph.D. for his suggestions, comments and insights. I would also like thank the Nord Pool AS for providing me access to their data.

Bachelor Thesis Proposal

Author	Štěpán Svoboda
Supervisor	doc. PhDr. Ladislav Krištoufek, Ph.D.
Proposed topic	Short-term electricity price forecasting - evaluation of selected hybrid models

---

Preliminary scope of work

The amount of renewable energy installed in Europe rose substantially in the last decade. This lead to the ever-growing intraday market to become more important over time. It is still far smaller and less important than the day-ahead market but the rise in its importance is unlikely to stop. Thus the intraday forecasting in the electricity market is becoming more and more important as the transactions within the electricity spot-market can take place sometimes even less than an hour before the delivery. The goal of this thesis is to compare an interesting hybrid model to a proven benchmark model.

There has been a substantial amount of literature written in the last decade or so. Ever since the electricity markets got deregulated this kind of research started to matter and have a practical application. Weron (2014) has written a very interesting summary about all of the state-of-art models and methods used not just in short-term but also in medium- and long-term forecasting. Álvarez et al. (2011) proposed in his paper the Pattern Sequence based forecasting and Che & Wang (2010) have proposes in their article the hybrid SVRARIMA model. There are dozens of papers written on this topic and they combine all possible machine learning and econometric models. They differ whether they use more data than just the previous prices (e.g.: Neupane 2013) or not (e.g.: Álvarez 2011, Che & Wang 2010) and whether they use single method or an ensemble learning method (e.g.: Shen et al. 2013) or whether they use some kind of data preprocessing and optimization algorithm (e.g.: Zhang et al. 2012)

The main contribution of my work should lie in comparison of SVRARIMA model to PSF model. The hybrid model combining Supporting Vector Machines and Autoregressive Integrated Moving Average model has been used before but was never directly compared to one of the more favorite benchmark models, which is the PSF.

The existing SVRARIMA model will be used to forecast the electricity prices from the NordPool area and these results will be compared to the benchmark PSF model using root mean squared error and mean absolute (percentage) error.

Outline

1. Introduction + motivation
2. Methodology
   (a) Previous studies
   (b) Data
   (c) Models
3. Results
   (a) Comparison of models
   (b) Discussion of results
4. Conclusion

Core bibliography

1. Che, Jinxing, & Wang, Jianzhou. 2010. Short-term electricity prices forecasting based on support vector regression and Auto-regressive integrated moving average modeling. *Energy Conversion and Management*, 51(10), 1911-1917

2. Alvarez, Francisco Martinez, Troncoso, Alicia, Riquelme, Jose C, & Ruiz, Jesus S Aguilar. 2011. Energy time series forecasting based

on pattern sequence similarity. *IEEE Transactions on Knowledge and Data Engineering*, 23(8), 1230-1243

3. Zhang, Jinliang, Tan, Zhongfu, & Yang, Shuxia. 2012. Day-ahead electricity price forecasting by a new hybrid method. *Computers & Industrial Engineering*, 63(3), 695-701

4. Weron, Rafał. 2014. Electricity price forecasting: A review of the stateof- the-art with a look into the future. *International journal of forecasting*, 30(4), 1030-1081

5. Jin, Cheng Hao, Pok, Gouchol, Paik, Incheon, & Ryu, Keun Ho. 2015. Short-term electricity load and price forecasting based on clustering and next symbol prediction. *IEEJ Transactions on Electrical and Electronic Engineering*, 10(2), 175-180

6. Shen, Wen, Babushkin, Vahan, Aung, Zeyar, & Woon, Wei Lee. 2013. An ensemble model for day-ahead electricity demand time series forecasting. *Pages 51-62 of: Proceedings of the fourth international conference on Future energy systems.* ACM

7. Ziel, Florian, Steinert, Rick, & Husmann, Sven. 2015. Efficient modeling and forecasting of electricity spot prices. *Energy Economics*, 47, 98-111

8. Shayeghi, Hossein, Ghasemi, A, Moradzadeh, Mohammad, & Nooshyar, M. 2015. Day-ahead electricity price forecasting using WPT, GMI and modified LSSVM-based S-OLABC algorithm. *Soft Computing*, 1-17

9. Neupane, Bijay. 2013. *Ensemble Learning-based Electricity Price Forecasting for Smart Grid Deployment.* Ph.D. thesis, Masdar Institute of Science and Technology

10. Liu, Heping, & Shi, Jing. 2013. Applying ARMA-GARCH approaches to forecasting short-term electricity prices. *Energy Economics*, 37, 152-166

Contents

1 Introduction

Since the last decade of the 20th century deregulation of the electricity markets has taken place all over the world. The sector that historically used to be a government-controlled monopoly has changed completely and became a standard market that is no longer only vertically or horizontally integrated. Many countries now possess a standard wholesale market for electricity where every market participant can buy or sell the desired amount. But electricity is still a long-way from being a standard commodity due to many of its special attributes, such as zero shelf life, possible negative price and the necessity of matching supply and demand in real time. These characteristics make the electricity market unlike any other and cause the daily, weekly and annual seasonality and generally unanticipated price spikes (Weron, 2014).

The main goal of this thesis is to compare the performance of two models in the intra-day market – the SVRARIMA (Support Vector Regression and Autoregressive Integrated Moving Average) model and the Pattern Sequence-based Forecasting model. They are also compared to the building blocks of the SVRARIMA model, the SVR and ARIMA models. These two models have never been compared before and neither have been applied on the intra-day market price. The SVRARIMA model have been used for forecasting in the day-ahead market (Che & Wang, 2010) and PSF have been utilized for day-ahead forecasts as well (Alvarez *et al.*, 2011).

This thesis is divided as follows, in chapter 2 the electricity price forecasting is described with special focus on the electricity market general features, dynamics and its specifics, the forecasting horizon and the forecasting methods, their usage, strengths, weaknesses and the horizon for which they are used. Because no market is the same, the specifics of the Nord Pool market are described in the next chapter as the empirical analysis is done on its data. An extensive survey of the previously used methods is carried out in the next chapter as well to show the wide variety of existing approaches. Chapter 3 describes the hybrid SVRARIMA model and focuses on the logic behind the combination of SVR and ARIMA in a single model and also on the mathematical foundations of this approach. Chapter 4 serves the same purpose as chapter 3 in describing the logic and fundamentals behind the Pattern Sequence-based Forecasting model. In chapter 5 the exact data used are described, the practical aspects of forecasting are discussed and the performance of both models in the intra-day market is compared with the results of both models presented. The performance of both SVRARIMA building blocks is used as a benchmark. The last chapter summarizes the thesis and stresses the most important thoughts and results of this thesis.

2 Electricity price forecasting

Electricity price forecasting (EPF) is a very specific discipline for a number of reasons. In this chapter the unusual characteristics of EPF will be described alongside the electricity market itself. The structure of this chapter is as follows - the description of the electricity market, the forecasting horizon and the division of the types of methods used in the forecasting.

2.1 Description of the electricity market

The electricity market consists of four parts - the generation capacity, the transmission and distribution network and the retail supply of the electricity. The deregulation of the power market that started in the 90's is concerned with making the generation and retail competitive. Both the transmission and distribution networks are considered a natural monopoly (Bajpai & Singh, 2004) and remain either in the hands of the states or under constant regulatory supervision. The two competitive parts are then the wholesale and the retail market. In this thesis only the wholesale market is of interest to us and from now on *electricity market* refers to the *wholesale market*.

Many aspects of the electricity market are highly unusual. The electricity behaves as a non-storable commodity, which has many consequences and makes it unlike any other commodity. The demand and supply need to be matched in real-time, which calls for the existence of Transmission Systems Operators (TSO's) that control and approve all the final outcomes of the physical contracts. The outcome of the necessity of real-time demand-supply matching is the increased volatility of electricity (Bajpai & Singh, 2004). Necessarily this feature of the electricity market only heightens the need for risk management and hedging done by various swaps, swaptions and contracts for differences, Benth & Koekebakker (2008), where also more detailed description of aforementioned financial instruments can be found. Another difference from a standard commodity is the transportation of electricity. There has to be a viable physical link (i.e.: a transmission network with the correct frequency) and that is widely limiting in terms of the market size. Another downside is the loss of electricity during transmission and distribution that is always present and depends on the quality of the network itself. All of the above mentioned reasons are why the electricity is sometimes called a *flow* commodity (Benth & Koekebakker, 2008).

The specifics of the electricity market, that have just been described are the cause for many special properties of the electricity prices and are behind the many difficulties of the EPF. The prevalence of price spikes is one of the troublesome features that make the forecasting complicated. Due to the volatile nature of electricity demand and supply, especially lately with the enormous rise in renewable electricity generation capacities, the price spikes are quite frequent (Carmona & Coulon, 2014) and very hard to predict (Weron, 2014). Another feature is the existence of negative prices that is becoming more common with the growing renewable capacity and renders many forecasting approaches commonly used in financial markets useless (Carmona & Coulon, 2014). The last peculiarity is the seasonality of the electricity prices. Yearly, monthly and daily seasonality are the outcome of very specific electricity demand curve. There are many palpable causes for this seasonality such as weather, national holidays, working days etc. This seasonality naturally has an influence on the occurence of the spikes as well (e.g. the weather seasonality, which influences the renewable sources and also enlarges the demand).

The electricity price on the wholesale market is an ambiguous term as there are several types of price. The terms commonly used are: *real-time, intra-day, day-ahead* and *forward price* (Carmona & Coulon, 2014; Weron, 2014; Benth & Koekebakker, 2008). The main concept driving the electricity price is the merit order effect. The idea behind is that the sources with the lowest marginal price are dispatched as first and those with highest are dispatched as last - this succession is changed only due to reasons such as system reliability, transmission congestion and other similar reasons (Bajpai & Singh, 2004).

The real-time market is organized by the TSO and is used for upward or downward market regulation to deal with power deficit or surplus. In the case of Nord Pool market the real-time price is determined based on a Walrasian auction (Benth & Koekebakker, 2008). The price in the day-ahead market, in case of Nord Pool *Elspot*, is usually referred to as the *spot* price, according to Carmona & Coulon (2014) . Elspot is open the day prior to the electricity delivery and up to 36 hours ahead and is responsible for nearly 90% of the volume traded (EPEXSPOT, 2016) although the number is shrinking. The intra-day market, known as Elbas, is much smaller but due to the ever increasing renewable capacity is becoming more important over time (NordPool, 2017). This market is operational continuously and trades all the way up to one hour ahead of delivery. The rest of electricity is traded on the basis of longer term contracts and are of no interest to us in this thesis.

2.2 Forecasting horizon

There are three forecasting horizons that are commonly considered: the short-, medium- and long-term (Weron, 2014) forecasting. Unfortunately the literature does not provide us with the thresholds for the boundaries between the different horizons, though there are some approximate threshold values.

Under the term *short-term forecasting*, the generally understood time frame is anywhere from few minutes to few days ahead. This forecasting is mostly used for the day-ahead and intra-day trading and is utilized by the firms in the day-to-day market operations. *Medium-term forecasting* usually refers to anywhere from few days to few months ahead. This type of modeling is preferred for risk management, derivatives pricing and balance sheet calculations - this strain of forecasting is often focused on the distributions of prices rather than the individual point forecasts. *Long-term forecasting* refers to any time beyond the horizon of a few months. The main aim of this type of forecasting is long-term investment profitability analysis. For example a decision on power plant construction may be based on the long-term forecasting.

2.3 Forecast evaluation

There are many metrics used to evaluate and compare the evaluation methods of forecasting models and as of now there is no standard metric that is widely used in EPF setting (Weron, 2014). One of the popular measures is *absolute error*: $AE_h = |P_h - \hat{P}_h|$, where P_h is actual and $\hat{P}_h$ is predicted price. There are many variations of this measure as *daily/weakly mean absolute errors* with the MAE being counted on the basis of last 24 or 168 errors. The most popular measure is *absolute percentage error*: $APE_h = AE_h/P_h$ and there is also its adjusted version – *daily/weakly mean absolute percentage error* calculated from the last 24 or 168 errors. The issue with this measure is that when the prices are low the MAPE values are quite high and vice versa. The negative prices are another issue because of the interpretability of MAPE values in this setting.

$$RMSE_{(T=24\ or\ 168)} = \sqrt{\frac{1}{T}\sum_{h=1}^{T}(P_h - \hat{P}_h)^2} \quad (1)$$

Root mean square error is another popular measure that is often used. The quadratic nature of this measure makes it very susceptible to spikes and this can be partially rectified by altering the measure to *root mean square percentage error* or its weekly or daily version.

In this thesis a statistical test for predictive accuracy is also used in evaluating the different methods. The Diebold-Mariano test originally created by Diebold & Mariano (1995) and then later modified by Harvey *et al.* (1997) is a statistical test with null hypothesis of no difference between the two compared forecasts. The advantage of this test is its generality as the loss function can be non-quadratic, asymmetric and the errors can be non-Gaussian, nonzero mean and serially or contemporaneously correlated. This allows us to test virtually any two methods and to get a valid statistical result in the end. The test itself is not described in greater detail in this thesis but in case of interest details can be found in Diebold & Mariano (1995);

Harvey *et al.* (1997). We will restrict ourselves here to mentioning that the final DM statistic of the modified Diebold-Mariano test follows the standard normal distribution and thus the critical regions for two-sided test are for the DM statistic when $DM < -1.96$ or $DM > 1.96$ given the standard significance level of 5%. The construction of the statistic implies values in lower critical region mean the superior performance of the original forecasting method and values in upper critical region the superior performance of the alternative method.

2.4 Forecasting approaches

The literature concerned with the EPF is already quite rich and there are several very good reviews of the existing approaches – Carmona & Coulon (2014); Weron (2014); Benth & Koekebakker (2008); Hu *et al.* (2009); Aggarwal *et al.* (2009); Martínez-Álvarez *et al.* (2015). Weron (2014) is probably the most extensive and complex one. Weron's division of the existing approaches is used here instead of the older method divisions made by Aggarwal *et al.* (2009) - Game Theory, Time Series and Simulation models or Hu *et al.* (2009) - Data Mining, Time Series and Simulation models. The division of the forecasting approaches has no generally accepted form and the choice of the division is based on author's preference and judgment.

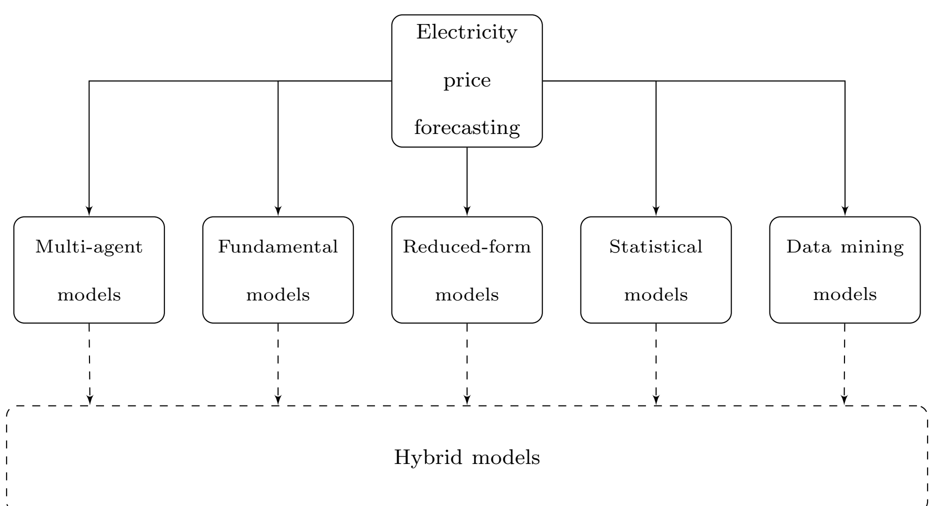


Figure 1: Based on Weron (2014); Martínez-Álvarez *et al.* (2015)

Many of the models presented below are quite hard to classify as they are some version of a hybrid model - using one or several methods from one group and another method from a different group. The author classifies them based on his judgment and the information found in several wonderful reviews, surveys and complex papers mentioned in the beginning of this section.

2.4.1 Multi-agent models

This type of models builds on the approach that existed before the deregulation of the electricity market in the 90's - the so-called *production-cost models* (PCM). The models used to forecast the price by matching the demand and supply by stacking up the generation units - without the presence of strategic bidding this approach was valid and appropriate. However with the deregulation and the loss of stable structure these models were not applicable anymore and new models based on them were created (Weron, 2014). These approaches can be roughly divided into two subgroups - *equilibrium (game theoretic)* and *agent-based models*, that became more popular lately. In this sense the equilibrium models are a step above the PCMs due to introduction of strategic behavior, while the agent-based simulation models are capable of describing properties of electricity markets indescribable by the static equilibrium (Weron, 2014).

Ventosa *et al.* (2005) created a very thorough survey and named the popular approaches. Two of those are applicable in the case of EPF, equilibrium and simulation models, and Weron (2014) based his classification of *multi-agent models* on this article. "Equilibrium models represent the overall market behavior taking into consideration competition among all participants" and "simulation models are an alternative to equilibrium models when the problem under consideration is too complex to be addressed within a formal equilibrium framework" write Ventosa *et al.* (2005) in their description of these models. Ventosa *et al.* further divide the equilibrium models into Cournout equilibrium and Supply function equilibrium models. The simulation techniques are divided into agent-based and equilibrium models. These subgroups are not further described in this thesis as they are beyond the scope of this work.

Let us take a look at one example of a hybrid multi-agent model. Kintsakis *et al.* (2015) use an agent-based model, which demonstrates beautifully the adaptability and flexibility of this class of models. Kintsakis *et al.* use a combination of signal-preprocessing and a machine learning method - Linear Wavelet Neural Network - in tandem with a genetic optimization algorithm - Particle Swarm Optimization (PSO) - to describe the behavior of market participants.

This group of models and forecasting techniques has one significant advantage - its flexibility. These models, especially the agent-based ones, are an amazingly flexible instrument for the analysis of strategic decision making. Unfortunately this strength is also its greatest weakness as the flexibility requires many conditions and assumptions to be properly chosen and defended within the simulation framework. Another downside is the need for definition of all the market participants and all the ways in which they interact, which is quite cumbersome given the strategic behavior of all the participants. This can be seen on the classification of market participants as sellers or buyers in the case of a generation company or a distribution company as they can be both. The possible strategies are thus nontrivial.

For all of these reasons the multi-agent models are best suited to answer questions about qualitative issues as they lack in the quantitative department, i.e. forecasting the electricity prices with high precision.

2.4.2 Fundamental models

The *fundamental models* aim to predict the electricity price by properly describing the physical and economic relationships in electricity production. The relationship between the price and the factors behind it is the alpha and omega of this approach. The methods used to predict the price itself can be various (e.g. statistical, data mining or hybrid models). Two classes of these models can be identified - the parsimonious and parameter-rich models.

The latter group is described on the case of Vehviläinen & Pyykkönen (2005). Vehviläinen & Pyykkönen tailor their model to the Nordic market due to its unique structure of high hydro power plant share. They consider factors from climate (e.g. temperature, snow-pack changes or water inflow) supply and demand. Overall they use 27 parameters and 29 mathematically defined relationships of the fundamental factors. Their model is able to capture the price movements accurately on monthly scale, which falls in the medium-term forecasting horizon.

The parsimonious models are much simpler than the parameter-rich models. There are many different ways these models can be developed and on which processes are based. In this thesis this method is demonstrated on the case of Carmona *et al.* (2013). Carmona *et al.* use stochastic models of the *bid stack* (inverse supply curve) to determine the spot price from the load and fuel prices. The structure of the bid stack allows for an easy and quick changes in the merit order curve in the future. Some of the benefits of this approach are quite obvious, while some not so much. The forecasted price naturally reflects the fundamental factors that are responsible for the demand and supply, which counts as an obvious advantage, while the closed-form formulas, which are of great use in derivative pricing belong to the not so obvious advantages.

The last examples discussed here are clear representatives of hybrid models. The first one is Kristiansen (2012) as an example of a statistical (standard auto regressive) model with exogenous (fundamental) variables. Kristiansen uses Nord Pool data and as the exogenous variables uses Nordic electricity demand and Danish wind power. In the second example Liebl *et al.* (2013) base their approach on the merit order model and reduced-form models (described in the next section). Liebl *et al.* use a two step approach - firstly they model the relationship between the spot price and the demand on daily basis and secondly they parametrize the obtained series of functions using a functional factor model. They have employed a parsimonious model approach but other factors can be included than just the load. The last example is an interesting thesis from Neupane (2013), where quite complex ensemble-learning model is employed. Neupane uses an algorithm combining three machine learning methods (random forests, neural networks and support vector machines), which has a fixed and varying weights version and bases its predictions on the performance of the individual methods. The features used in the forecasting are selected by the algorithm from a large data set including many different variables such as various weather and calendar data. Neupane's results are compared to state-of-art method (PSF) and his complex approach comes out on top.

The fundamental models have two obvious weaknesses. The first one is data. The amount of data needed for the fundamental models, especially the parameter-rich models, is sizeable and not all of it is easily accessible. Another issue is the frequency of data gathering, monthly or weekly intervals are not an exception in this case. Parsimonious models are less vulnerable to this but they still usually use daily data. That makes these models suitable for medium-term forecasting and thus risk management and derivative pricing. The fundamental models excel in the same area as the *reduced-form models* and the survey from Carmona & Coulon (2014) does a great job as an introduction to the fundamental models. They usually lack the precision to focus on hourly data. The second weakness are the stochastic fluctuations influencing the fundamental variables. The price predictions are very susceptible to the relationships between the fundamentals and a slight misspecification can severely hamper the capabilities of the model.

2.4.3 Reduced-form models

The *reduced-form models*, sometimes called *stochastic* or *quantitative*, are specific in what they try to forecast. These models do not try to forecast the hourly prices in the future, they aim at describing the main features of the electricity prices on daily basis, such as the marginal distributions in the future, price dynamics and relationships between commodity prices (Weron, 2014). This type of models is in direct competition with the aforementioned fundamental models as both types specialize and excel at describing the medium-term horizon and play major role in derivative pricing and hedging (i.e. risk management). Naturally if the process chosen to describe the electricity price is ill suited the results are disappointing and if the model is not simple and fast enough then it cannot fulfill its role in the trading departments (Weron, 2006). According to Weron (2014) these models hold the best of two worlds - they are based on methods developed to model other similar commodities or interest rates and include also actuarial and econometric features - they are a middle ground between models that are well suited to describe the unique nature of electricity prices and simplicity that makes them usable in trading departments.

The reduced-form models can focus either on the spot prices or on the forward prices. The latter belongs to the domain of mathematical finance rather than EPF and is thus beyond the scope of this thesis. The models focused on the spot price can be further divided into two groups, the *jump-diffusion* and *Markov-regime switching models*. The jump-diffusion models are usually special cases of general stochastic differential equation for the difference of spot electricity price. This price needs to be detrended and deseasonalized. Naturally there are many different versions of these models but that differentiation is beyond the scope of this thesis.

The Markov-regime switching models (MRS) rectify one major weakness of the jump-diffusion models and that is the inability to properly model consecutive spikes (Weron, 2014). The main idea behind these models is modeling the spot price process using several different regimes with different underlying stochastic processes. The transition from one regime to another is assumed to be an unobserved Markov chain R_t. More detailed description of Markov-regime switching models is beyond the scope of this thesis and more information regarding these models can be found in Liebl *et al.* (2013).

2.4.4 Statistical models

The *statistical*, sometimes also called *econometric, models* are based on predicting the current price using the previous prices and/or exogenous factors such as climate-, supply- and/or demand-related variables. The first versions of these EPF models were just adjusted load models, where load was replaced by price. They are primarily used in short-term forecasting, e.g. day-ahead market, but also intra-day market. Another asset of these models is their interpretation - it is possible to use the components to explain the electricity price, to attach meaning to the individual parts of the equation. The critique of these models is usually aimed at their limited capability of modeling non-linear behavior of the electricity price, even though in empirical comparisons their performance does measure up to their non-linear competitors (Weron, 2014).

This group of models can be divided into several subgroups - *similar-day methods, regression models, AR-, ARX-, TAR- and GARCH-type models*. Sometimes also the *VAR-type models* can be utilized in this setting, which are a natural multivariate extension of the AR-type models (Martínez-Álvarez *et al.*, 2015). The similar-day (or *naive*) method is quite popular technique in EPF. It is usually used as a benchmark method to test other methods against it. The idea is quite simple - Tuesday is similar to last week's Tuesday. According to Nogales *et al.* (2002) models that are not carefully calibrated do fail this basic test quite regularly. The regression models can quite strongly overlap with the ARX- type of models as many of them have lagged electricity price regressors but are still called regression models. The standard model is usually fitted using least squares methods:

$$P_t = \mathbf{B}_t \mathbf{X}_t + \varepsilon_t, \tag{2}$$

where $\mathbf{B}_t$ is $1 \times k$ vector of constant or time-varying coefficients (in case of *time-varying regression*) and $\mathbf{X}_t$ is $k \times 1$ vector of regressors with ε_t being a standard error term.

The AR-type of models are approaches called either ARMA(p, q) - *Auto-*

regressive Moving Average or ARIMA(p, d, q) - *Auto-regressive Integrated Moving Average models*. ARIMA models are a generalization of the ARMA models and originate in the work of Box & Jenkins (1976). The ARMA models consist of an AR (p) and an MA (q) processes and assume that the described time series is stationary or at least weakly stationary. If that is not the case then the time series is differenced and we obtain an ARIMA model. That means an ARIMA $(p, 0, q)$ is identical to an ARMA(p, q) model:

$$\phi(B) \bigtriangledown^d X_t = \theta(B)\varepsilon_t, \tag{3}$$

where B is the backward shift operator, where $B^h X_t \equiv X_{t-h}$ $\phi(B)$ represents $\phi(B) = 1 - \phi_1 B - ... - \phi_p B^p$ and $\theta(B)$ represents $\theta(B) = 1 - \theta_1 B - ... - \theta_q B^q$ with $\phi_1 ... \phi_p$ and $\theta_1 ... \theta_q$ representing the actual coefficients of the autoregressive and moving average processes. These models also allow for existence of seasonality within the SARIMA framework. ARIMA(p, d, q) $\times(P, D, Q)_s$, where the $(P, D, Q)_s$ describes the seasonal part of the ARIMA process with s signaling the number of observations within one period (e.g. 24 hours for daily period, seven days for a weekly period etc.)

The VAR-type models are used in multivariate setting, e.g. one model for each hour of the day instead of one model for all hours in the day. An example for $N = 2$ of VAR(1) is:

$$y_{1,t} = \alpha_{1,1}\, y_{1,t-1} + \alpha_{1,2}\, y_{2,t-1} + \varepsilon_{1,t} \tag{4}$$

$$y_{2,t} = \alpha_{2,1}\, y_{1,t-1} + \alpha_{2,2}\, y_{2,t-1} + \varepsilon_{2,t}. \tag{5}$$

This can be easily generalized to the number of lags (i.e. p) and number of specific time series (i.e. N) necessary.

The ARX-type of models is the same as the AR-type with one exception - inclusion of the *eXogenous* variables, which makes them also fundamental and thus also hybrid. Just like before ARMAX and ARIMAX (or ARIMA-E) models can be created based on the same logic as before with inclusion of the fundamental variables.

The *threshold autoregressive models* (i.e. TAR models) are a *regime-switching method* just like the MRS models. The difference being certainty of the switch between the regimes, i.e. an observable variable decides when the change between regimes happens. These models, called TAR models, based on the work of Tong & Lim (1980), are not described as they are beyond the scope of this thesis.

The last type of model in this group is the *Generalized Autoregressive Conditional Heteroskedastic* or GARCH (p, q) model introduced by Bollerslev (1986). The AR(X)-type of model assumes homoskedasticity, which is often unachievable in the case of financial and other time series and the GARCH models address this issue. Their more detailed description is beyond the scope of this thesis.

Liu & Shi (2013) use an ARMA-GARCH model, which is unlike ordinary GARCH model quite competitive. Bordignon *et al.* (2013) use several different forecasting techniques - an ARMAX model, a standard linear regression model, a time-varying parameter regression model and Markov-regime switching model. His work is focused on the forecast combination and his conclusion is quite straightforward - in 99% of cases combination of forecasts leads to better results than one single model, given ex ante comparisons. Choosing one model is riskier than choosing combination of several as every model excels in some circumstances but also lacks in other.

Ziel *et al.* (2015) use an econometric approach and apply a periodic VAR-TARCH model with inclusion of wind and solar power alongside load as explanatory variables. The estimation technique used is a *least absolute shrinkage and selection operator* also known as lasso (description of this technique is beyond the scope of this thesis). Ziel (2016) is another paper that uses the lasso estimation technique based on the prices in the same hour up to 36 days in the past. The lasso approach is used to avoid overfitting. The idea of this model builds on the work of Weron & Misiorek (2008).

The first example of a hybrid statistical model is an already mentioned fundamental hybrid model, i.e.: Kristiansen (2012), a typical ARX-type model. The next example combines wavevelet transform, ARIMA model and Radial Basis Function Neural Network (RBFN) with PSO. Shafie-Khah *et al.* (2011) forecast the electricity price by (i) wavelet transformation (ii) ARIMA (iii) RBFN to catch the non-linear patterns in the ARIMA residuals (iv) PSO to optimize RBFN structure.

The biggest weakness of these models is the occurence of spikes. The statistical models do not perform well in presence of non-linearity and the spikes are the prime example of this weakness. Janczura *et al.* (2013) is a great review and study of this phenomenon. The proposed solutions can be summarized as (i) outlier detection (ii) spike replacement. Many methods have been proposed such as recursive filters and wavelet filtering for detection and chosen threshold or mean of neighboring prices as the replacement values. This list is non-exhaustive and aims to be only illustrative.

2.4.5 Data mining models

This class of techniques is very rich and wide. The term *data mining* is meant to refer to everything from machine learning and artificial intelligence to evolutionary models and data transformation. The techniques encompassed in this classification are *neural networks, support vector machines, genetic programming, fuzzy logic*, various types of *clustering, data transformation and ensemble methods.* Most of hybrid models use one or more methods from the list above and combine them usually with a statistical method. Weron (2014) referred to this group as *computational intelligence models* as there is no standard division of the forecasting approaches. These models are used mainly for short-term electricity forecasting as their biggest strength is their ability to capture the non-linearity involved in the electricity price.

Neural networks are probably the most widely used machine learning technique in the EPF. Neural networks consist of several layers of neurons with each neuron being either activated or not on the arrival and potentially sending the information further along, just like the human nervous system. This is only a very basic description but a more in-depth and mathematical description are beyond the scope of this thesis as well as methods based on neural networks. Haykin & Lippmann (1994) is a good description and introduction in this problematic.

Support vector machines are another machine learning technique. It is described in detail in the next chapter and here is only the basic idea presented. This technique takes data that are linearly inseparable in the current dimension and transforms them into a higher dimension, where they can be separated by a hyperplane. Vapnik (2013) is the founding text of this methodology and Cristianini & Shawe-Taylor (2000) is a very nice introduction and learning material.

Genetic programming is an extremely wide class of techniques. They are stochastic processes based on phenomena observed in nature. Their goal is to imitate the evolutionary process and use the ”strongest exemplars”, i.e. the best performing cases from past generations. These methods are most often utilized in fine-tuning the parameters of other data mining techniques. More detailed discussion does not take place here as it is beyond the scope of this thesis due to its complexity.

Fuzzy logic is another tool often utilized by other methods, whether it is neural networks, clustering or even statistical methods (e.g. standard OLS regression). This logic, unlike the classical Boolean logic, does not have only 1 and 0 (i.e. true and false) as its values, but can attain other values often based on qualitative ranges (e.g. medium, peak and off-peak electricity prices).

Clustering is also used only as a tool or as a part of other methods. The most commonly used clustering method is k-means clustering. It is an unsupervised learning method, where the parameters (number of clusters) must be specified beforehand. The Patter Sequence-based forecasting algorithm is based on this method and is described in more detail in later chapter as well as k-means clustering.

Data transformation techniques are applied before and after the modeling. There are several types of these transformations, the most common are wavelet-type transformations. There is a number of other transformations, such as discrete cosine transformation, but these methods are also beyond the scope of this thesis.

The last type of data mining methods described here is the *ensemble methods*. They take several predictive algorithms and by combining their results they obtain the final prediction. Several of these methods are described below. Probably the most well known one is the random forest - an ensemble of decision trees. More detailed description is beyond the scope of this thesis.

Three papers utilizing the neural networks are presented here. Anbazhagan & Kumarappan (2014) created a model that uses a standard feed-forward neural network on data that were transformed by the discrete cosine transformation. Shrivastava & Panigrahi (2014) use a version of neural networks known as extreme learning machines in combination with the wavelet transformation. The last NN paper if from Chaâbane (2014a), who proposes a model that combines an ARFIMA model (generalized version of ARIMA) and neural networks as forecasting engine with the ARFIMA model being focused on the linear part and the neural network forecasting the non-linear part.

Chaâbane (2014b) uses another version of his previous model with one difference - the neural network is replaced with least square support vector machines (LSSVM). Unfortunately he does not compare these two methods with each other, only with their building blocks and an ARIMA-ANN model. Both methods outperform their building blocks and the chosen benchmark model. Kaya *et al.* (2015) forecast the electricity price with a support vector regression (SVR) and wavelet transformation model. Che & Wang (2010) propose a SVRARIMA model, that is described in detail in the next chapter. Zhang *et al.* (2012) create a model based on Che & Wang (2010), a hybrid LSSVM and ARIMA model, with the data being transformed by the wavelet transformation and the LSSVM parameters optimized by the PSO algorithm. The last paper based on the SVM method is Shayeghi *et al.* (2015)

that pre-processes the data using generalized mutual information and wavelet packet transformation, then forecasting by Bayesian version of LSSVM (LSSVM-B) with a version of artificial bee colony as an optimization algorithm.

Interestingly the SVM methods have been found to be as accurate as the neural networks but more consistent, meaning they need less time for training (Sansom *et al.*, 2003), due to its property of global optimum existence (more on this in next chapter).

Two more models are introduced now, that are not based on either of the two most popular techniques. The PSF model by Alvarez *et al.* (2011) is described in more detail later in the thesis. The second model is from Jin *et al.* (2015) and utilizes clustering like the PSF model. It uses prediction by partial matching method on the cluster labels of the clustered data to forecast the next symbol.

The last group of techniques is the ensemble methods. One was already mentioned in the fundamental models section - Neupane (2013). He uses in his thesis an ensemble method of SVM, random forest and neural networks and an optimization algorithm to predict the electricity price using the fundamental data. The second is an ensemble PSF method from Shen *et al.* (2013). They use 5 PSF models based on different types of clustering methods - k-means, hierarchical, k-medoids, fuzzy C-means and self-organizing map. Then an iterative prediction procedure is used to optimize the result.

The biggest strength of this class of models is also their biggest weakness. The ability to capture the complex and non-linear behavior can lead very easily to over-fitting and poor out-of-sample performance. Another downside is the comparison of these models - they are incomparable unless calibration and testing data sets are completely identical.

3 Description of the SVRARIMA model

This model was first introduced by Che & Wang (2010) and this chapter describes the model they have created. Unless otherwise stated the paper from Che & Wang (2010) is the source for this chapter.

The idea of combination of two vastly different models (i.e. *hybrid* model) to achieve better results is quite old based on the paper from Granger (1989). The selection of the models that are to be combined is not discussed in the paper from Che & Wang (2010) or the one from Granger (1989) either and this discussion is beyond the scope of this thesis.

The foundation of this model is the combination of two effective methods. While SVR is very good at capturing the nonlinear patterns the ARIMA model excels with linear patterns. The SVR is employed as first and the ARIMA model is used to regress the residuals. The justification for this order is the ε-insensitivity of the SVR model and the overall spikiness and nonlinear patterns dominance in the electricity price.

The problem is defined by Che & Wang in such a way: time serial $\{X(t)(t = 1, 2, ..., n)\}$, reconstruction phase space $\{Y(j)\}$, time-delays τ and embedded dimension $m(j = 1, 2, ...N, N = n - (m - 1))$, with the phase space being:

$$Y(j) = [X(j), X(j - \tau), X(j - 2\tau), ..., X(j - (m - 1)\tau)]. \tag{6}$$

Then utilizing the Takens embedding theorem we get a mapping $F : R^m \to R^m$, which makes $Y(j + \tau) = F(Y(j))$. Then there is a smooth map in the reconstructed phase space $\bar{F} : R^m \to R$ such that:

$$X(j + \tau) = \bar{F}(X(j), X(j - \tau), X(j - 2\tau), ..., X(j - (m - 1)\tau)). \tag{7}$$

This means that the time series prediction can be performed given the existence of concrete function expression. The SVR can be used to obtain this function.

3.1 Support Vector Regression

Support Vector Machines by Vapnik (2013) are classification and regression techniques, that optimize its structure based on the input data. Due to the complexity of this method only a brief and concise summary is given in this thesis. Some very extensive summaries that go beyond the scope of this thesis are Cristianini & Shawe-Taylor (2000); Smola & Schölkopf (2004); Burges (1998).

For training data $(x_1, y_1), ..., (x_n, y_n)$, where x_i are the vectors with input values and y_i the appropriate output values for x_i, the ε-insensitive SVR aims to find a function $f(x)$, that has the deviation from the target y_i at most ε at all times, while being as "flat" as possible. This problem can be written down as an optimization problem:

$$\underset{\omega,b,\xi,\xi^*}{min} \quad \frac{1}{2}\omega^T\omega + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{8}$$

$$\text{subject to} \begin{cases} y_i - (\langle\omega, x_i\rangle + b) \leq \varepsilon + \xi_i \\ (\langle\omega, x_i\rangle + b) - y_i \leq \varepsilon + \xi_i^* \\ \xi, \xi^* \geq 0 \end{cases} \tag{9}$$

where n stands for the number of samples, ξ_i represents the upper training error and ξ_i^* the lower training error subject to the ε-insensitive tube $|y_i - (\langle\omega, x_i\rangle + b)|$. The last term, C, represents the regularized constant, that determines the trade-off between the regularization term and the empirical error, also $C > 0$. The SVR creates the $f(x)$ such that: (i) ξ_i and ξ_i^* are minimized to achieve the minimal training error and (ii) to make the function "flat" and penalize too complex functions we minimize $\frac{1}{2}\omega^T\omega$. The final decision function has the form of:

$$f(x, \alpha_i, \alpha_i^*) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)K(x, x_i) + b \tag{10}$$

and can be found by utilizing the properties Lagrange multipliers, Kernel trick and the optimality constraints. The Lagrange multipliers and Kernel trick are described below.

3.1.1 Lagrange multipliers

In Equation 10 α_i, α_i^* represent the Lagrange multipliers. They can be obtained by maximizing the dual function of Equation 8 and then take the following form:

$$\begin{aligned} \underset{\alpha_i,\alpha_i^*}{max} \quad & \sum_{i=1}^{n} y_i(\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^{n} y_i(\alpha_i + \alpha_i^*) \\ & -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)\langle x_i, x_j\rangle \end{aligned} \tag{11}$$

$$\text{subject to} \begin{cases} \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0 \\ 0 \leq \alpha_i \leq C \\ 0 \leq \alpha_i^* \leq C \\ i = 1, 2, ..., n \end{cases} . \tag{12}$$

In Equation 10, only some $(\alpha_i - \alpha_i^*)$ are not equal to zero, which comes from the Karush-Kuhn-Tucker's conditions of solving a quadratic programming problem. The *support vector* itself refers to the approximation error of data point on non-zero coefficient equal or larger than ε. Because errors lower than ε are acceptable, the data from the training set inside the "ε-tube" do not contribute to the cost nor the solution of the problem.

3.1.2 Kernel trick

The key to non-linear extension of the SVR is the Equation 11 and the existence of the so-called *kernel trick*. The dot product of $\langle x_i, x_j\rangle$ from the Equation 11 becomes a kernel function $\langle\phi(x_i), \phi(x_j)\rangle = K(x_i, x_j)$ in the case of non-linearity. The function $\phi : R_d \rightarrow F$ presents the idea of mapping the input space into a feature space with a higher dimension. The kernel function used in this thesis is called the radial basis kernel and its mapping space has infinite number of dimensions.

$$K(x, x') = exp(-\sigma||x - x'||^2) \tag{13}$$

3.2 Autoregressive Integrated Moving Average

The ARIMA model is in detail described in the Section 2.4.4, Equation 3. This model has three parameters (p, d, q), that refer to number of autoregressive terms, differences and moving average terms. Three stages are present during the modeling – model identification, parameter estimation and diagnostic checking. First the model is selected based on Akaike Information Criteria (AIC):

$$AIC = 2k - 2ln(\hat{L}), \tag{14}$$

where k is the number of parameters and $\hat{L}$ is the maximized likelihood function. The function is defined in the following way: $\hat{L} = P(x|\hat{\theta}, M)$, where $\hat{\theta}$ are the parameter values, which maximize the function and M symbolizes the model itself. Then the parameters are estimated and at last the diagnostics are done using residuals plots and auto-correlation functions.

Lastly it is important to say that any significant non-linearity hinders the ARIMA model and modeling the linear patterns does not leave the non-linear patterns unchanged.

3.3 Forecasting

As is discussed above, the hybrid model SVRARIMA is meant to capture both the non-linear and linear parts of the electricity price, with which non-hybrid models might have problems. The modeling of one part inadvertently changes the other but the change can be kept small, thanks to the capability of the SVR model to keep the linear patterns quite intact in comparison with other methods such as neural networks.

As is customary with SVM the data are standardized before the modeling begins.

$$X(t) = \frac{X'(t) - \bar{X}}{\hat{\sigma}} \tag{15}$$

where $X(t)$ is the standardized price, $X'(t)$ is the original price, μ is the average of all prices within the time series and σ is the standard deviation of said time series.

The hybrid model forecasts the electricity price Y_t based on the idea, that the electricity price has a linear and a non-linear component:

$$Y_t = N_t + L_t \tag{16}$$

where N_t stands for the non-linear part of electricity price and L_t for the linear part. Both parts are estimated from the data set using a simple approach, first the SVR is used to model the N_t and the residual at time t is then denoted ε_t.

$$\varepsilon_t = Y_t - \hat{N}_t, \tag{17}$$

$\hat{N}_t$ stands for the forecasted value from SVR at time t.

The residuals from the SVR model, i.e. ε_t, are then forecasted using the ARIMA model in the following fashion:

$$\varepsilon_t = f(\varepsilon_{t-1}, \varepsilon_{t-2}, ..., \varepsilon_{t-n}) + e_t \tag{18}$$

where f stands for the linear function of the ARIMA model and e_t is the residual from the second step, i.e. ARIMA model.

4 Description of the PSF model

This model is based on the work of Alvarez *et al.* (2011) and unless stated otherwise that paper is the source for this chapter.

The main idea of this technique is to assign each day into a cluster and then operate just with the cluster label instead of the entire feature set. In the case of daily prices prediction it would mean that we operate with a single label per day instead of a 24-dimensional point describing electricity price in a given day.

The PSF algorithm has two phases – the clustering and the forecasting phase. Both are described below according to the paper from Alvarez *et al.* and this thesis does not go into greater depth than the original paper. The works on which the original paper is based are not discussed here and can be found in the original paper.

4.1 Clustering

The first step of clustering that must be present is the data transformation. The original paper has chosen the following transformation:

$$x'_j = \frac{x_j}{\frac{1}{N}\sum_{i=1}^{N} x_i} \tag{19}$$

where N represents the number of hours in a day, x'_j is the price after the transformation and x_j is the price for the j-th hour of the day.

The clustering has to assign each day to one of the K clusters. In this case the clustering is quite complex as every day is one point in a 24-dimensional space. This step allows us to greatly decrease the complexity of prediction as we have only one dimension after the clustering and that is the labels itself.

Two questions must be addressed before the clustering can take place - which technique should be used and how many clusters should there be? The existing literature does not give us a definitive answer and thus the authors based on the evidence in previous work regarding these kinds of data sets have chosen the K-means algorithm as the optimal choice. Other algorithms can be used as was seen in Shen *et al.* (2013). In this thesis the methodology of the original paper is followed.

With the choice of K-means algorithm we have partially answered also our second question. The number of clusters must be chosen a priori. This is a very difficult task and once again there is no one best way to choose the number. The solution that Alvarez *et al.* have come up with is to use the three most widely used indexes – Silhouette index, Davies-Bouldin (DB) index and Dunn (DU) index. If majority of the indexes have the same ideal K, then that one is chosen. Otherwise the second best place is also considered and the best index is once again chosen based on the highest representation within the set.

4.1.1 The Silhouette index

The *silhouette function* gives us a measure of the quality of the clustering partition done. The value *a(i)* represents the average distance of i from cluster A to all the objects in A with the average distance of i to all objects in the in $C \neq A$ is called $d(i, C)$. Then the $d(i, C)$ ic calculated for all clusters and the minimal one is chosen in following way:

$$b(i) = \min_{C \neq A} d(i, C) \;\; with \; i \in A. \tag{20}$$

The equation above describes the dissimilarity of the object i to the nearest neighboring cluster, the overall index value can then be denoted as follows:

$$silh(i) = \frac{a(i) - b(i)}{max\{a(i), b(i)\}} \tag{21}$$

and the values of this function can range from $+1$ to -1. These values represent an appropriate and inappropriate cluster respectively. The value 0 represents information, that the object i can be in both clusters with the same similarity properties. By taking the average of $silh(i)$ over the number of objects to be classified and maximizing this value, the optimal number of clusters can be found.

4.1.2 The Davies-Bouldin index

Th Davies-Bouldin index is based on identifying compact clusters far from each other. This index is defined as:

$$DB_K = \frac{1}{K}\sum_{i=1}^{K} \underset{j \neq i}{max} f_{i,j}, \tag{22}$$

where K represents the number of clusters and $f_{i,j}$ stands for:

$$f_{i,j} = \frac{diam(C_i) + diam(C_j)}{d(C_i, C_j)}, \tag{23}$$

where $d(C_i, C_j)$ stands for the dissimilarity between C_i and C_j in such a way:

$$d(C_i, C_j) = \underset{\substack{x \in C_i \\ y \in C_j}}{min} ||x - y||, \tag{24}$$

where $|| \cdot ||$ stands for the norm and $diam(C_i)$ stands for:

$$diam(C_i) = \left(\frac{1}{n_i}\sum_{x \in C_i} ||x - z_i||^2\right)^{\frac{1}{2}}, \tag{25}$$

with n_i being the number of points and z_i the centroid of i-th cluster – C_i. Low values of the Davies-Bouldin index are optimal, as the optimal number of clusters is found when this index is minimized for the data set.

4.1.3 The Dunn index

The Dunn index aims to find the clusters, which have low intracluster and high intercluster distance. It is defined in the following way:

$$DU_K = \underset{i}{min} \; \underset{j \neq i}{min} \; f_{i,j}, \tag{26}$$

where

$$f_{i,j} = \frac{d(C_i, C_j)}{max_m \ diam(C_m)}. \tag{27}$$

The $d(C_i, C_j)$ stands for the same as in the Davies-Bouldin index and the $diam(C)$, the intracluster distance function, is defined in following way:

$$diam(C) = \max_{x,y \in C} ||x - y||. \tag{28}$$

This optimal number of clusters is found when the Dunn index reaches the maximum value – compact and nicely separated clusters.

4.2 The PSF algorithm

The aim of this approach is to use the prices from all days up to $d-1$ and then forecast the 24 prices of the day d. The vector $X(i) = [x_1, x_2, ..., x_{24}]$ represents the 24 hourly electricity prices on the day i and thus $X(i) \in \mathbb{R}^{24}$.

Then $L_i \in \{1, ..., K\}$ is the label of the day i based on the clustering described above with K representing the number of clusters. These labels can be turned into a following sequence:

$$S_W^i = [L_{i-W+1}, L_{i-W+2}, ..., L_{i-1}, L_i], \tag{29}$$

where W represents the length of the window. The determination of this parameter is done later in this chapter.

The PSF algorithm first searches for sequences exactly equal to S_W^{d-1}.These sequences are then denoted as ES_d, which can be noted as follows:

$$ES_d = \{j \ such \ that \ S_W^j = S_W^{d-1}\}. \tag{30}$$

If no such sequences are found then the length of the window is lowered by 1, which can be done successively until W is equal to 1. The predicted values of electricity price are then achieved by averaging the prices of the days following the found subsequences, i.e. ES_d. It can be written down as

$$\hat{X}(d) = \frac{1}{size(ES_d)} \sum_{j \in ES_d} X(j+1), \tag{31}$$

where size(ES_d) is the volume of elements in the set ES_d.

In case of predicting multiple samples or a medium- or a long-term forecast, after each prediction the data must be updated, the clustering must be repeated and the prediction step must be done once again.

4.3 Window size determination

Before the prediction can take place the parameter W must be determined on a training set. This can be expressed mathematically as:

$$\sum_{d \in TS} ||\hat{X}(d) - X(d)||, \tag{32}$$

where $\hat{X}(d)$ is the forecasted price for day d and $X(d)$ is the actual price for day d. TS then refers to the training set previously mentioned. The parameter W cannot be obtained using the standard mathematical programming methods and must be determined empirically by using *cross-validation*.

The approach from the original paper is followed here and the n-fold cross-validation is used. The n-fold cross-validation splits the data set into n subsets and then each data set is used as a validation data set for the model trained on the $n-1$ remaining subsets. The n results are then combined to give us the parameters of the final model.

Alvarez *et al.* used 12-fold cross-validation with one fold for every month. In every fold are these forecasting errors determined. Then the errors are expressed as $e_{month}\{W = j\}$ for $j = 1, ..., W_{max}$, where the authors have determined the value of W_{max} to be 10. The final errors for each window are then determined as follows:

$$e_j = \frac{1}{n} \sum_{i=1}^{n} e_{month}\{W = j\}, \tag{33}$$

where $month = \{Jan, ..., Dec\}$ and $n = 12$.

Then the W is selected, which minimizes the average error of the evaluated folds.

$$W = arg\ min\{e_j\}\ with\ j = 1, ..., W_{max}. \tag{34}$$

5 Empirical application

In this chapter the data used in this analysis, practical specifications of the methods and the software used to forecast the electricity prices are described. Afterwards the results of the forecasting are presented and discussed.

5.1 Data

The data used have been provided to the author by Nord Pool. Their intra-day hourly prices from years 2015 and 2016 are used by all of the methods and the data from 2014 are used on top of that by the PSF method for the initial training period. The intra-day market trading takes place in the Nordic and Baltic countries and Germany. The market itself is described in previous chapters and details are not discussed here.

The behavior of electricity price and its peculiarity is discussed in earlier chapters and deserves a reminder. The electricity price suffers not only from non-stationarity and seasonality (e.g.: daily, monthly, yearly) but also from non-zero mean, high volatility, presence of spikes and negative or null prices. All of these problems are addressed within the modeling done below.

5.2 Forecasting

The practical aspects of forecasting, like the software used and the methodology, are discussed now. The rolling window approach is utilized to obtain enough results for a robust comparison of the models. This approach gives us a 100 weekly metrics, which allow us to make a qualified comparison and not be subject to a random noise and/or patterns within the data itself.

First the ARIMA model is discussed, then SVR, SVRARIMA and PSF models are discussed as well. The software used in all instances is R, version 3.3.2, and RStudio, version 1.0.136. The specific packages used are also mentioned.

For the ARIMA modeling the *forecast R package version 8.0* (Hyndman, 2017) is used to implement the (S)ARIMA framework and forecast the prices. The rolling window approach was utilized in following way – based on the existing literature (e.g.: Zhang *et al.* (2012); Che & Wang (2010)), the author's judgment and customs the train to test ratio is three to one and for each one-week testing set three-week training set is used. The first test week is the fourth week of 2015 and the first three weeks of 2015 are the first training sets. There are 100 iterations, where every iteration moves the test and train period by one-week into the future, of this setup, which results in 100 weekly metrics to evaluate this model with its competitors. As was previously discussed no external regressors are used.

The rolling window approach for SVR is very similar to the one described above with one difference - the feature selection. The number of lags of first differences used to forecast the electricity price is chosen by the author to perform a basic sensitivity analysis. The modeling of the price is done using two, three, four and five lags of the price at time t. The best performing feature set is then chosen.

After the feature set is determined the parameter tuning comes in place. The best performing parameters are found during the tuning period using cross-validation and then by applying the final model to our test set. The *mlr R package version 2.11* (Bischl *et al.*, 2016) is used as a unifying interface for the tuning, feature selection and model training.

The SVRARIMA forecasting was done by applying the ARIMA model to the residuals from the SVR models.

The approach for the PSF technique is slightly different due to the dissimilarity of the methods. The testing weeks are the same as previously to allow fair comparison but the training set is different. The year 2014 and the first three weeks of 2015 are used as the basic training set and then after every iteration the training set is prolonged by one week. At every iteration the clustering and determination of the window size must be done. The *PSF package version 0.4* (Bokde *et al.*, 2016) is used for the forecasting and parameter tuning.

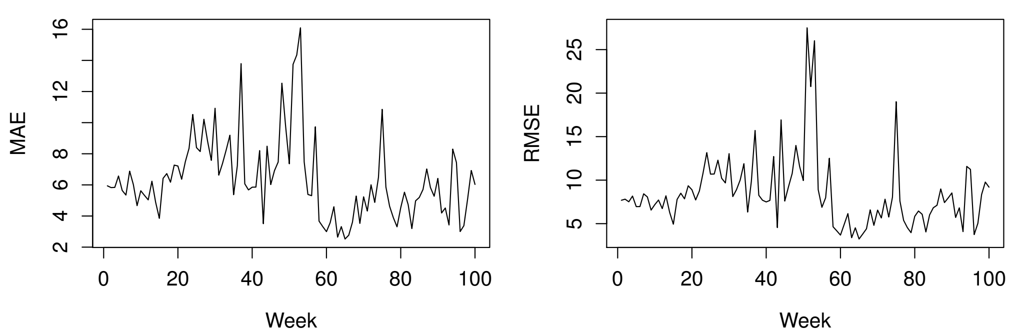
5.3 Results and discussion

The obtained results from the modeling are presented and discussed in the following paragraphs. The comparison of the performance is done using the metrics presented in the section 2.3. Due to the rolling approach window our results are not very susceptible to outliers and can be described as robust.

Firstly the individual models and their performance are presented, then a more in-depth comparison of models is done with focus on the explanation of the results. The Diebold-Mariano test for predictive accuracy is used in the more detailed description to compare the models among each other. This is followed by a final discussion of the overall results and findings of this thesis.

5.3.1 Results

The first model presented here is the PSF model. The metrics used for the forecast evaluation are the MAE and RMSE. The performance of this model turned out to be very unsatisfactory. The reasons for this outcome are discussed after the presentation of the performance of other models.



(a) Weekly MAE

(b) Weekly RMSE

Figure 2: Weekly average statistics obtained from the Pattern Sequence-based Forecasting model for every week in our rolling window test period

In the case of SVR model a basic sensitivity analysis was done by making 4 models with 2, 3, 4 and 5 lagged prices as the explanatory variables (i.e. features). When the price at time t was forecasted, the values $t-1, t-$

$2, ..., t-5$ were used in the forecasting.

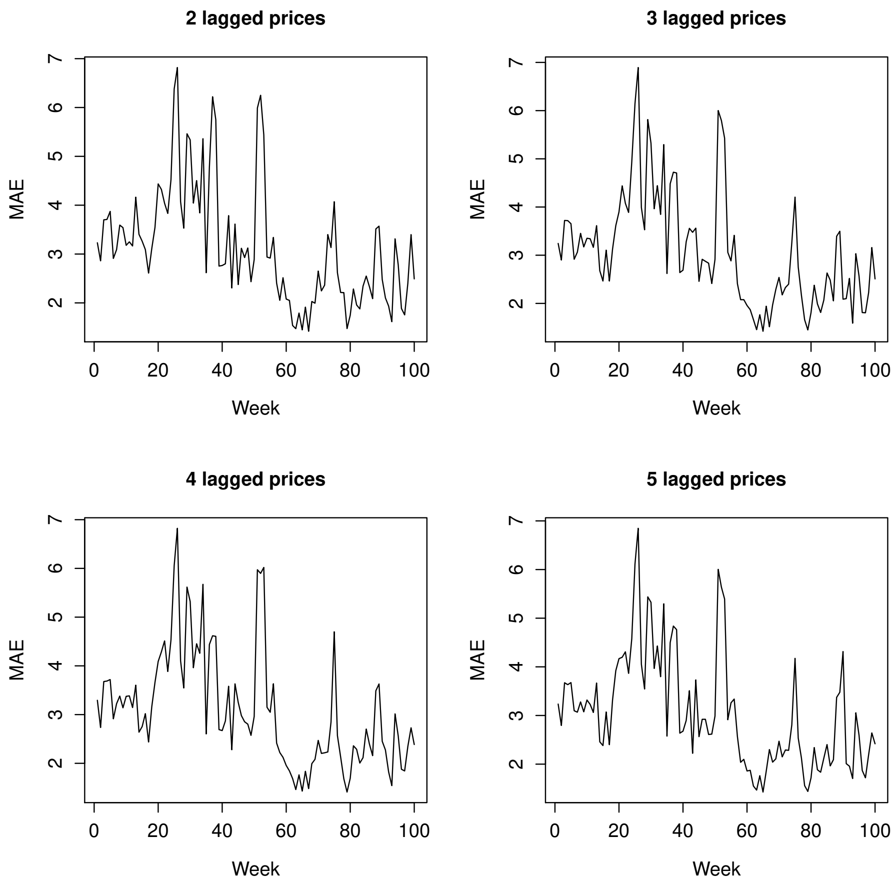


Figure 3: Weekly average MAE obtained from the Support Vector Regression model for every week in our rolling window test period for all considered feature sets

The performance of the SVR model measured by the MAE and the RMSE measure can be found in Figure 3 and 4. The cases of 4 and 5 previous prices used seem to be the best but a more detailed comparison is carried out later after all of the models' performances are presented. The results in this case are drastically better than the performance of the hybrid PSF model.

Quick look at the results here also explains to us why the RMSE measure alone is not very popular in the EPF setting as was discussed in the second chapter. Thanks to the spikes the results can be quite volatile. As we can see on the Figure 4 the RMSE measure amplifies these spikes and the entire performance measure can be thrown off balance.

The Figure 5 presents the results of the well-known ARIMA model. The

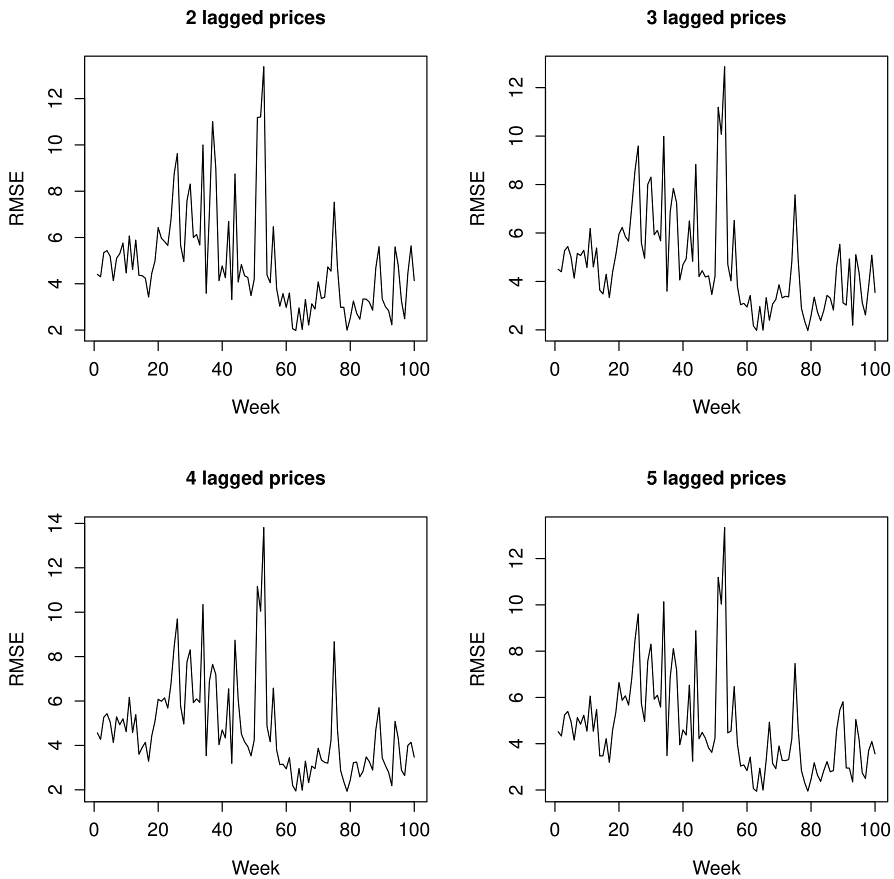
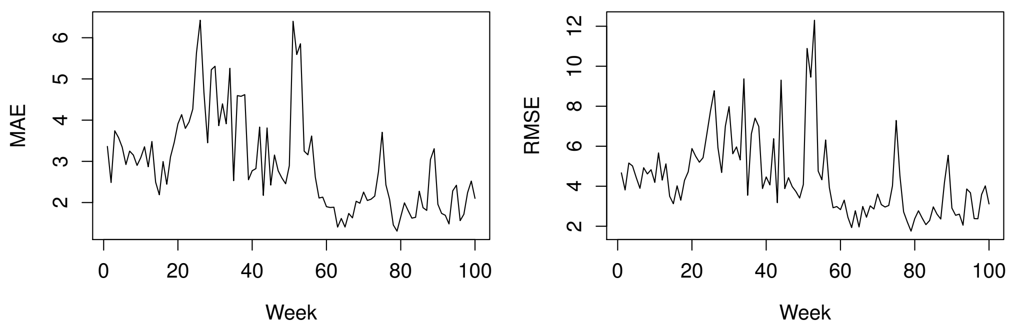


Figure 4: Weekly average RMSE obtained from the Support Vector Regression model for every week in our rolling window test period for all considered feature sets

first look at the chart seems to suggest similar level of performance in the case of SVR and ARIMA model. A more detailed discussion and comparison follows but a simple look does not give us more information unlike in the case of PSF model.

The SVRARIMA model and its performance measures follow in the Figure 6 and 7. The SVRARIMA model seems to be very similar in performance to both already mentioned models and no clear difference between the three methods is easily observable. More detailed comparison and results follow later.



(a) Weekly MAE

(b) Weekly RMSE

Figure 5: Weekly average statistics obtained from the ARIMA model for every week in our rolling window test period

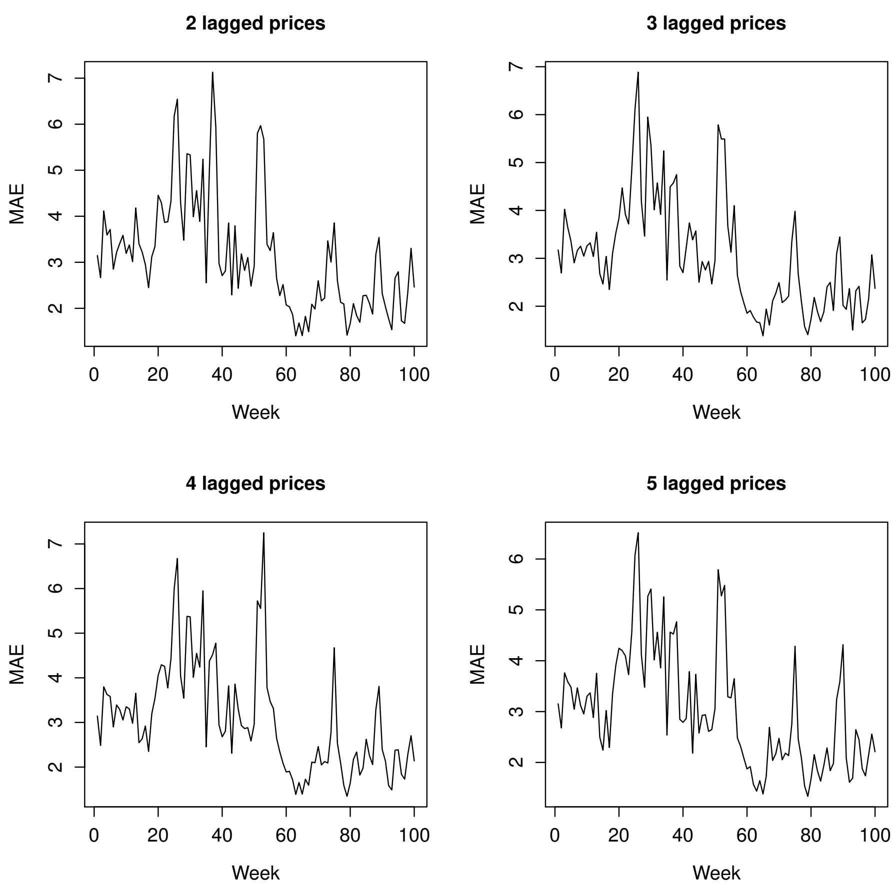


Figure 6: Weekly average MAE obtained from the SVRARIMA model for every week in our rolling window test period for all considered feature sets

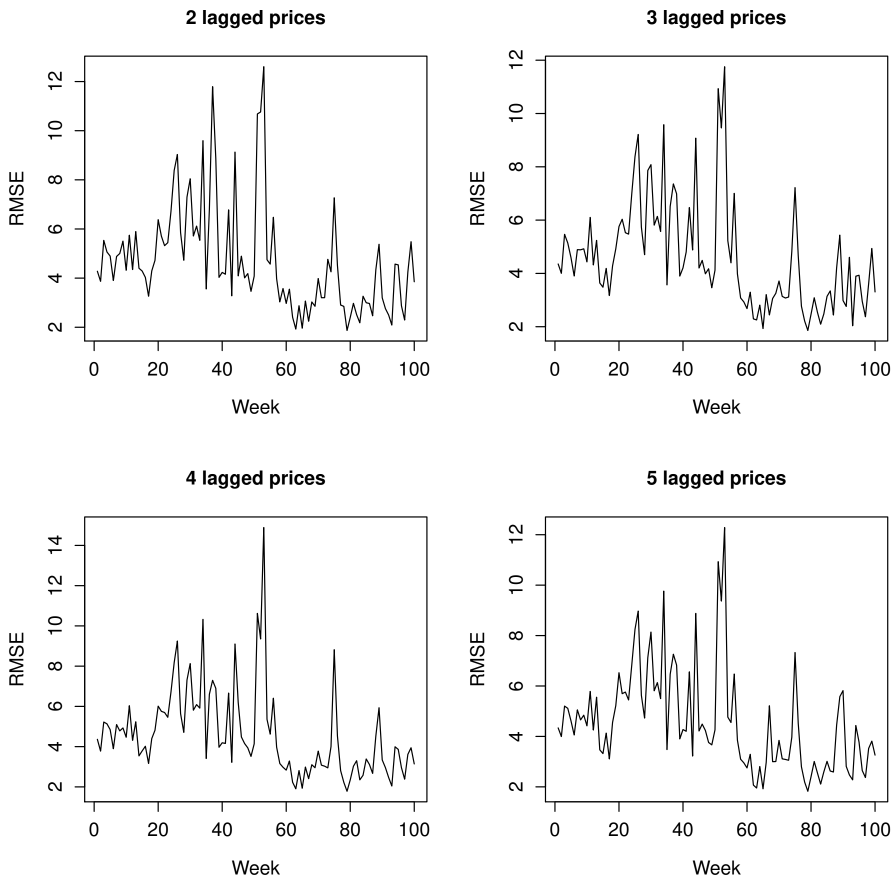


Figure 7: Weekly average RMSE obtained from the SVRARIMA model for every week in our rolling window test period for all considered feature sets

The Table 1 with all the results aggregated gives us a good idea about the actual performance of each model. Both the MAE and RMSE measures are the sum of the corresponding errors for all the iterations, i.e. sum of 100 weekly measures. Thus a simple average MAE or RMSE for the 100 weeks can be obtained by a simple division.

The first observation is quite obviously the bad performance of the PSF model. The PSF model is ill-suited for this kind of forecasting most likely due to its incapability of hour-to-hour forecasting. This model can forecast only a day ahead and thus is quite severely outperformed by the models, which are able to perform on the hour-to-hour basis. This model might be (and according to literature is) well-suited for short-term forecasting but not intra-day forecasting as it was not created for this kind of modeling.

Table 1: Performance measures - the sum of all 100 weekly performance statistics

	MAE	RMSE
PSF	635.9026	848.0119
ARIMA	294.2193	442.3174
SVR - 2 lags	316.4178	487.4136
SVR - 3 lags	308.8589	476.5567
SVR - 4 lags	309.6362	475.2557
SVR - 5 lags	306.2951	473.1802
SVRARIMA - 2 lags	314.0151	471.4788
SVRARIMA - 3 lags	305.5741	460.2413
SVRARIMA - 4 lags	306.76	461.4763
SVRARIMA - 5 lags	303.4642	456.8097

The Table 1 also confirms our original observation of some sort of equivalency between the ARIMA and the SVR model or a very slight dominance of the ARIMA model. The difference seems to be quite small in the case of MAE – on average a 0.12 to 0.15 difference per week when we do not consider the worst SVR model. The SVRARIMA model seems to steadily outperform the SVR model but the ARIMA model appears to outperform the SVRARIMA model. The difference is even smaller than in the case of SVR – ARIMA comparison. The difference is studied later and the statistical significance of our results is also tested.

5.3.2 An in-depth comparison of forecasts

Several ways of comparing the forecasts are used now. Firstly the equivalence or slight dominance of the ARIMA model over the SVR(ARIMA) model is more closely studied. The difference between the weekly average metrics is plotted to see how often one method outperforms the other and how much. Then the density of the ARIMA's and SVR(ARIMA)'s best version of weekly average MAE is plotted and compared to see the differences between the models better.

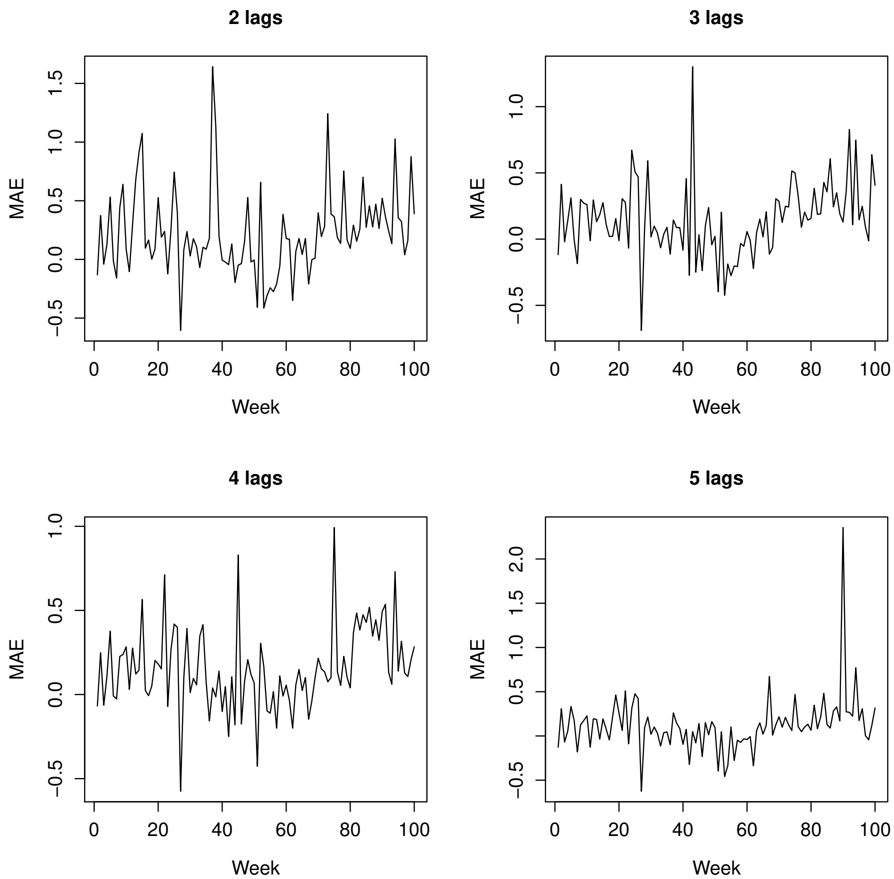


Figure 8: The difference between the weekly MAE statistic for the SVR and the ARIMA model – negative values imply SVR's better performance

The difference is plotted as $MAE_{ARIMA} - MAE_{SVR}$ and $MAE_{ARIMA} - MAE_{SVRARIMA}$. This means that positive values imply better performance of the ARIMA model and vice versa.

From the Figure 8 our initial observation is once again confirmed that the SVR and ARIMA models have similar performance but ARIMA seems to be slightly better overall than SVR.

The Figure 9 tells us a very similar story as the Figure 8. The ARIMA model overall appears to slightly outperform both our data-mining and hybrid model. Though we must not forget this assumption is not backed by any test of statistical significance and we should wait with our assessment of the models after we carry out a full analysis.

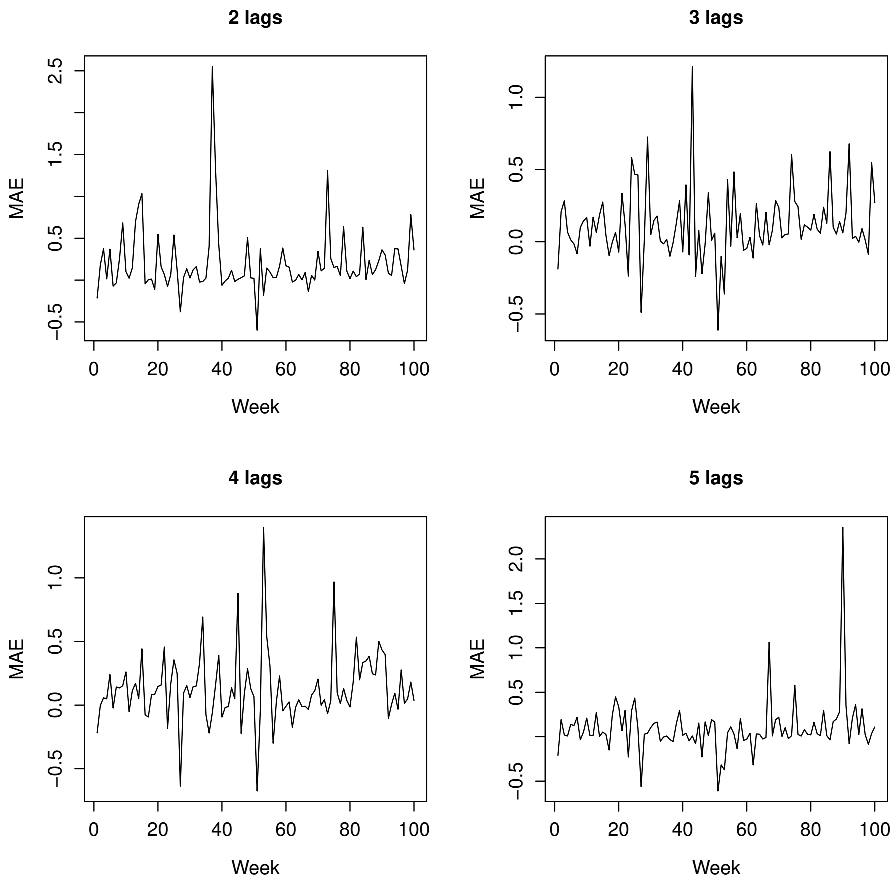


Figure 9: The difference between the weekly MAE statistic for the SVRARIMA and the ARIMA model – negative values imply SVRARIMA's better performance

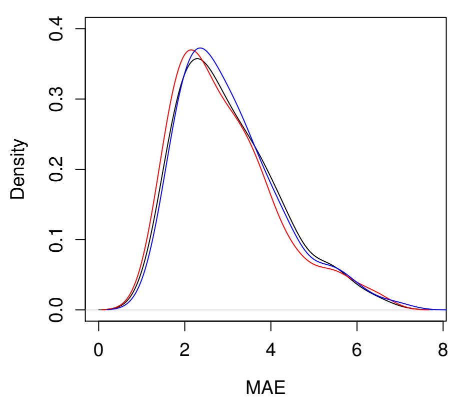


Figure 10: Each curve represents the theoretical density of the MAE distribution based on the observed values – black - SVRARIMA, red - ARIMA, blue - SVR

Figure 10 shows us the densities of the weekly MAE of the best model versions of SVR and SVRARIMA and of the ARIMA model.The comparison of the SVR and ARIMA model reiterates the previous conclusion and that the ARIMA model seems to be either marginally better or equivalent in forecasting accuracy. The SVRARIMA model once again appears to be very marginally better than SVR model and slightly worse than ARIMA model. This picture supports our initial assessment of the models.

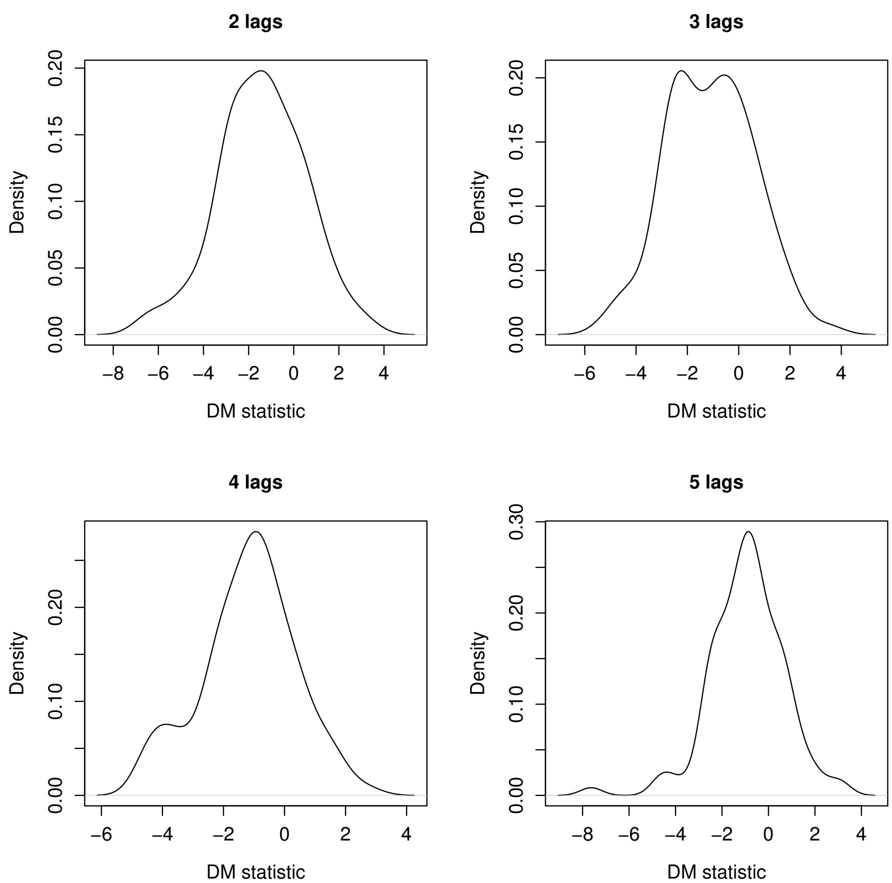


Figure 11: The density of the DM statistic from the comparison of the ARIMA and the SVR model. The statistic follows the standard normal distribution and in the regions beyond -1.96 and 1.96 we have the critical regions for null hypothesis rejection – both models have the same prediction accuracy.

Now the Diebold-Mariano test is carried out in three cases – ARIMA to SVR, ARIMA to SVRARIMA and SVR to SVRARIMA comparison. In the last case the appropriate lags are used as the test values for achieving objectivity and comparability. On the Figures 11, 12 and 13 we can see the density of the actual DM statistics in the cases described. The test is done for all the lags in cases of SVR and SVRARIMA to ARIMA comparison and in the case of SVR to SVRARIMA comparison the DM statistics are computed for the same basic SVR model, i.e. 2 lags to 2 lags comparison.

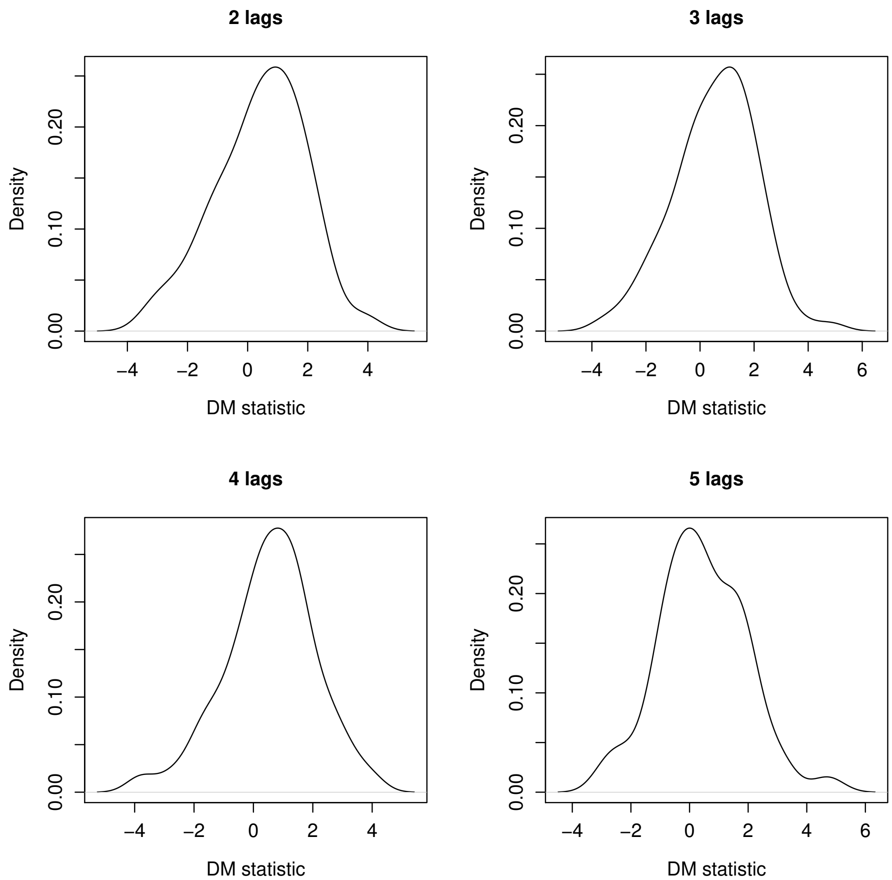


Figure 12: The density of the DM statistic from the comparison of the SVR and the SVRARIMA model. The statistic follows the standard normal distribution and in the regions beyond -1.96 and 1.96 we have the critical regions for null hypothesis rejection – both models have the same prediction accuracy.

The Figure 11 compares the SVR model to the ARIMA model. We can see the biggest number of the studied cases in the region between the two critical values of -1.96 and 1.96. This is confirmed by the statistics in the Table 2, which tell us the number of cases when SVR model is outperformed by ARIMA is between 22 and 38, while the number of cases, when the

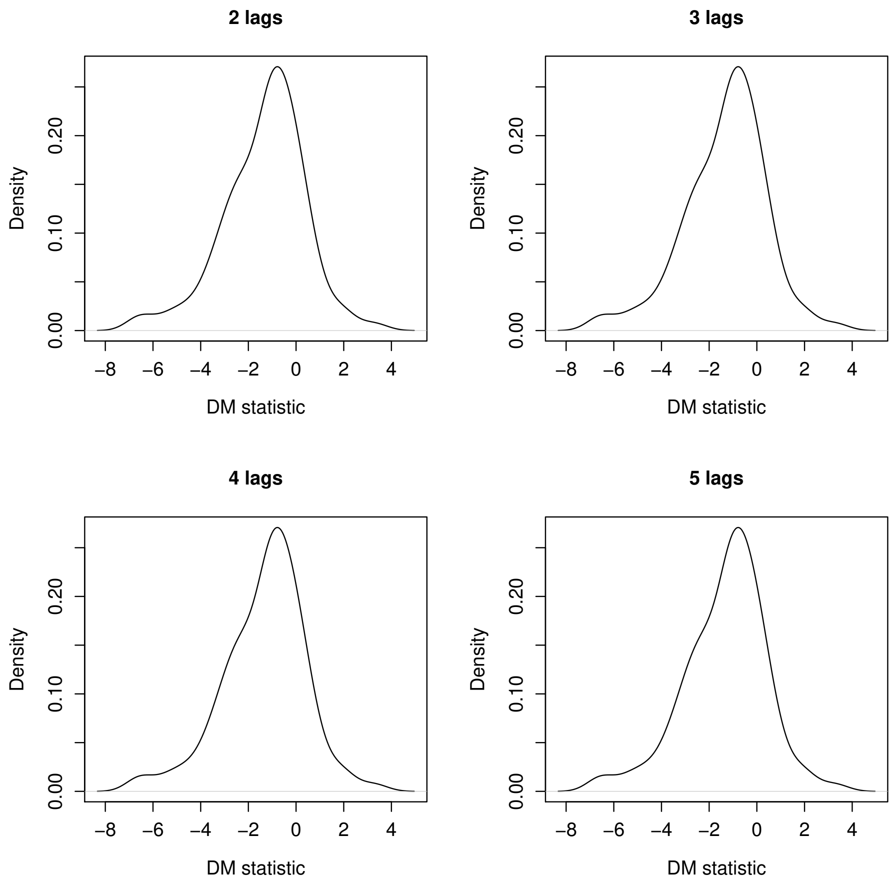


Figure 13: The density of the DM statistic from the comparison of the ARIMA and the SVRARIMA model. The statistic follows the standard normal distribution and in the regions beyond -1.96 and 1.96 we have the critical regions for null hypothesis rejection – both models have the same prediction accuracy.

outcome of the test is inconclusive is between 58 and 74. The ARIMA model is almost never statistically significantly outperformed by the SVR model. The evidence at our disposal clearly supports our conclusion of slight superiority of ARIMA model. This evidence is especially in case of the best SVR model somewhat weak and does not allow us an absolute decision. On average in our robust testing the SVR model was outperformed and the Diebold-Mariano test supports this hypothesis. The ARIMA model is not unequivocally better but the number of weeks when ARIMA dominated is far larger than in the case of SVR. Both are significantly lower than the instances when neither test has better results, which are statistically significant. This causes our inability to claim the ARIMA model more than just somewhat superior modeling technique in the case of EPF.

The Figures 12 and 13 lead us to a very similar conclusion. SVRARIMA model is slightly better than SVR, while ARIMA model outperforms SVRARIMA model. Overall these results are not extremely convincing as the cases in which neither model has outperformed the other at least at 5% confidence level is about three out of four times if we consider the best versions of the SVR and SVRARIMA models.

Table 2: Diebold-Mariano statistic – the number of weeks in each of the critical regions per method from left to right: alternative method worse, statistically not significant and original method worse, the alternative method is always written as second

	$DM < -1.96$	$DM > -1.96$ & $DM < 1.96$	$DM > 1.96$
ARIMA v SVR - 2 lags	38	58	4
ARIMA v SVR - 3 lags	35	62	3
ARIMA v SVR - 4 lags	26	73	1
ARIMA v SVR - 5 lags	22	74	4
SVR v SVRARIMA - 2 lags	8	78	14
SVR v SVRARIMA - 3 lags	6	79	15
SVR v SVRARIMA - 4 lags	5	79	16
SVR v SVRARIMA - 5 lags	6	79	15
ARIMA v SVRARIMA - 2 lags	33	65	2
ARIMA v SVRARIMA - 3 lags	31	35	4
ARIMA v SVRARIMA - 4 lags	29	68	3
ARIMA v SVRARIMA - 5 lags	19	76	5

5.3.3 Discussion

The results do come as quite a surprise considering the existing literature (Che & Wang, 2010). In this work both the SVRARIMA and SVR outperformed the basic ARIMA model but in our more robust comparison the other case seems to be true. Che & Wang (2010) used two chosen weeks when the non-linear nature of SVR appears to have prevailed so one explanation of our results may be the precise structure of the data at hand. Studying the precise details of the testing and training sets used in the cases when either SVR(ARIMA) or ARIMA were better at the 5% significance level can shine light on the differences and the explanation of different results from the existing literature. Future research in this area might help explain this.

We cannot forget that in the vast majority of cases the models did not have different predictive accuracy at 5% significance level. That suggests either some weeks have specific properties that make one model better than other (e.g. greater linear/non-linear part) or the better performance is due to randomness. This could be answered by conducting more investigation and ideally comparing multiple markets among each other to identify the roots of these differences. Each market can act quite differently from the other due to many local specifics such as renewable sources, consumption habits, weather patterns or the market development level. The difference may also be caused by the usage of intra-day market data, which brings another possibility of future research.

Another potential issue might have been the lack of fundamental variables in our models. The reasons for the choice of pure autoregressive models in this thesis are explained in previous chapters. Two examples of these reasons are the issues with obtaining the necessary data and the computational complexity. The data are mostly not that hard to find in the case of Nord Pool, but there are still problems, e.g. if we were to take air temperature into account we would have to model the prices for only some smaller areas of the Nord Pool market, because we cannot take the same temperature for Estonia and Norway or the northern and southern regions of Norway. The computational complexity would heighten considerably were the exogenous variables and their selection process included, especially in our case of rolling window analysis. The inclusion of such variables in our robust analysis case can be another area of future research.

6 Conclusion

The contribution of this thesis can be divided into two categories – the in-depth summary of the current electricity price forecasting literature with focus on the short-term forecasting and the evaluation and comparison of some statistical, data-mining and hybrid models.

The in-depth summary, which can be found in this thesis, can become a great source for anyone who wishes to learn more about this area of research as the literature review is quite broad and includes also a lot of context alongside some basic explanation of many methods and/or references to sources where this knowledge can be quite easily acquired. Several wonderful reviews and specialized surveys are also referenced within the literature review, which gives anyone the opportunity to immerse themselves in the topic using this thesis as a springboard.

Our comparison of the ARIMA, SVR, PSF and SVRARIMA models is quite reflective of them thanks to the robust nature of our modeling. It finds that the PSF model is wholly inadequate for this particular hour-to-hour forecasting setting. The SVR, SVRARIMA and ARIMA models are all on quite similar footing. The ARIMA model appears to be somewhat better than the rest but its dominance is not very strong. As we have stated more research is needed regarding these models, especially an investigation of this outcome's causes discussed at the end of previous section. The differences between different power exchanges, intra- and day-ahead market or the weeks themselves can all be the or one of the reasons for our outcome and their detailed study can provide us with answers. The comparison of the models used in this thesis and their versions with carefully selected fundamental variables is another potential area of research that might find if the impact of exogenous variables' inclusion in intra-day electricity price forecasting settings is statistically significant.

References

Aggarwal, Sanjeev Kumar, Saini, Lalit Mohan, & Kumar, Ashwani. 2009. Electricity price forecasting in deregulated markets: A review and evaluation. *International Journal of Electrical Power & Energy Systems*, **31**(1), 13–22.

Alvarez, Francisco Martinez, Troncoso, Alicia, Riquelme, Jose C, & Ruiz, Jesus S Aguilar. 2011. Energy time series forecasting based on pattern sequence similarity. *IEEE Transactions on Knowledge and Data Engineering*, **23**(8), 1230–1243.

Anbazhagan, S, & Kumarappan, N. 2014. Day-ahead deregulated electricity market price forecasting using neural network input featured by DCT. *Energy conversion and management*, **78**, 711–719.

Bajpai, Prabodh, & Singh, SN. 2004. Electricity trading in competitive power market: an overview and key issues. *In: International Conference On Power Systems, ICPS2004, Kathmandu, Nepal.*

Benth, Fred Espen, & Koekebakker, Steen. 2008. Stochastic modeling of financial electricity contracts. *Energy Economics*, **30**(3), 1116–1157.

Bischl, Bernd, Lang, Michel, Kotthoff, Lars, Schiffner, Julia, Richter, Jakob, Studerus, Erich, Casalicchio, Giuseppe, & Jones, Zachary M. 2016. mlr: Machine Learning in R. *Journal of Machine Learning Research*, **17**(170), 1–5.

Bokde, Neeraj, Asencio-Cortés, Gualberto, Martínez-Álvarez, Francisco, & Kulat, Kishore. 2016. PSF: Introduction to R Package for Pattern Sequence Based Forecasting Algorithm. *arXiv preprint arXiv:1606.05492*.

Bollerslev, Tim. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, **31**(3), 307–327.

Bordignon, Silvano, Bunn, Derek W, Lisi, Francesco, & Nan, Fany. 2013. Combining day-ahead forecasts for British electricity prices. *Energy Economics*, **35**, 88–103.

Box, George EP, & Jenkins, Gwilym M. 1976. *Time series analysis: forecasting and control, revised ed.* Holden-Day.

Burges, Christopher JC. 1998. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, **2**(2), 121–167.

Carmona, René, & Coulon, Michael. 2014. A survey of commodity markets and structural models for electricity prices. *Pages 41–83 of: Quantitative Energy Finance*. Springer.

Carmona, René, Coulon, Michael, & Schwarz, Daniel. 2013. Electricity price modeling and asset valuation: a multi-fuel structural approach. *Mathematics and Financial Economics*, **7**(2), 167–202.

Chaâbane, Najeh. 2014a. A hybrid ARFIMA and neural network model for electricity price prediction. *International Journal of Electrical Power & Energy Systems*, **55**, 187–194.

Chaâbane, Najeh. 2014b. A novel auto-regressive fractionally integrated moving average–least-squares support vector machine model for electricity spot prices prediction. *Journal of Applied Statistics*, **41**(3), 635–651.

Che, Jinxing, & Wang, Jianzhou. 2010. Short-term electricity prices forecasting based on support vector regression and Auto-regressive integrated moving average modeling. *Energy Conversion and Management*, **51**(10), 1911–1917.

Cristianini, Nello, & Shawe-Taylor, John. 2000. *An introduction to support vector machines.*

Diebold, Francis X, & Mariano, RS. 1995. Comparing forecast accuracy. *Journal of Business and.*

EPEXSPOT. 2016 (January). *Record volume on UK, overall Intraday markets.* Press release.

Granger, Clive WJ. 1989. Invited review combining forecastsâ"twenty years later. *Journal of Forecasting*, **8**(3), 167–173.

Harvey, David, Leybourne, Stephen, & Newbold, Paul. 1997. Testing the equality of prediction mean squared errors. *International Journal of forecasting*, **13**(2), 281–291.

Haykin, Simon, & Lippmann, Richard. 1994. Neural networks, a comprehensive foundation. *International journal of neural systems*, **5**(4), 363–364.

Hu, Linlin, Taylor, Gareth, Wan, Hai-Bin, & Irving, Malcolm. 2009. A review of short-term electricity price forecasting techniques in deregulated electricity markets. *Pages 1–5 of: Universities Power Engineering Conference (UPEC), 2009 Proceedings of the 44th International*. IEEE.

Hyndman, Rob J. 2017. *forecast: Forecasting functions for time series and linear models*. R package version 8.0.

Janczura, Joanna, Trück, Stefan, Weron, Rafał, & Wolff, Rodney C. 2013. Identifying spikes and seasonal components in electricity spot price data: A guide to robust modeling. *Energy Economics*, **38**, 96–110.

Jin, Cheng Hao, Pok, Gouchol, Paik, Incheon, & Ryu, Keun Ho. 2015. Short-term electricity load and price forecasting based on clustering and next symbol prediction. *IEEJ Transactions on Electrical and Electronic Engineering*, **10**(2), 175–180.

Kaya, Gamze Ogcu, Demirel, Omer Fahrettin, Zaim, Selim, & Gorener, Rifat. 2015. Day-ahead Electricity Price Forecasting Using a Wavelet-Support Vector Machine Conjunction Model. *Pages 179–186 of: Conference on Engineering And Technology Management GCETM-2015*.

Kintsakis, Athanassios M, Chrysopoulos, Antonios, & Mitkas, Pericles A. 2015. Agent-based short-term load and price forecasting using a parallel implementation of an adaptive PSO-trained local linear wavelet neural network. *Pages 1–5 of: European Energy Market (EEM), 2015 12th International Conference on the*. IEEE.

Kristiansen, Tarjei. 2012. Forecasting Nord Pool day-ahead prices with an autoregressive model. *Energy Policy*, **49**, 328–332.

Liebl, Dominik, *et al.* 2013. Modeling and forecasting electricity spot prices: A functional data perspective. *The Annals of Applied Statistics*, **7**(3), 1562–1592.

Liu, Heping, & Shi, Jing. 2013. Applying ARMA–GARCH approaches to forecasting short-term electricity prices. *Energy Economics*, **37**, 152–166.

Martínez-Álvarez, Francisco, Troncoso, Alicia, Asencio-Cortés, Gualberto, & Riquelme, José C. 2015. A survey on data mining techniques applied to electricity-related time series forecasting. *Energies*, **8**(11), 13162–13193.

Neupane, Bijay. 2013. *Ensemble Learning-based Electricity Price Forecasting for Smart Grid Deployment*. Ph.D. thesis, Masdar Institute of Science and Technology.

Nogales, Francisco Javier, Contreras, Javier, Conejo, Antonio J, & Espínola, Rosario. 2002. Forecasting next-day electricity prices by time series models. *IEEE Transactions on power systems*, **17**(2), 342–348.

NordPool. 2017. *The Power Market*.

Sansom, Damien C, Downs, Tom, Saha, Tapan K, *et al.* 2003. Evaluation of support vector machine based forecasting tool in electricity price forecasting for Australian national electricity market participants. *Journal of Electrical & Electronics Engineering, Australia*, **22**(3), 227–233.

Shafie-Khah, M, Moghaddam, M Parsa, & Sheikh-El-Eslami, MK. 2011. Price forecasting of day-ahead electricity markets using a hybrid forecast method. *Energy Conversion and Management*, **52**(5), 2165–2169.

Shayeghi, Hossein, Ghasemi, A, Moradzadeh, Mohammad, & Nooshyar, M. 2015. Day-ahead electricity price forecasting using WPT, GMI and modified LSSVM-based S-OLABC algorithm. *Soft Computing*, 1–17.

Shen, Wen, Babushkin, Vahan, Aung, Zeyar, & Woon, Wei Lee. 2013. An ensemble model for day-ahead electricity demand time series forecasting. *Pages 51–62 of: Proceedings of the fourth international conference on Future energy systems*. ACM.

Shrivastava, Nitin Anand, & Panigrahi, Bijaya Ketan. 2014. A hybrid wavelet-ELM based short term price forecasting for electricity markets. *International Journal of Electrical Power & Energy Systems*, **55**, 41–50.

Smola, Alex J, & Schölkopf, Bernhard. 2004. A tutorial on support vector regression. *Statistics and computing*, **14**(3), 199–222.

Tong, Howell, & Lim, Keng S. 1980. Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 245–292.

Vapnik, Vladimir. 2013. *The nature of statistical learning theory.* Springer science & business media.

Vehviläinen, Iivo, & Pyykkönen, Tuomas. 2005. Stochastic factor model for electricity spot priceâ"the case of the Nordic market. *Energy Economics*, **27**(2), 351–367.

Ventosa, Mariano, Baıllo, Alvaro, Ramos, Andrés, & Rivier, Michel. 2005. Electricity market modeling trends. *Energy policy*, **33**(7), 897–913.

Weron, Rafał. 2006. Modeling and Forecasting Electricity Loads. *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*, 67–100.

Weron, Rafał. 2014. 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.

Weron, Rafał, & Misiorek, Adam. 2008. Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models. *International journal of forecasting*, **24**(4), 744–763.

Zhang, Jinliang, Tan, Zhongfu, & Yang, Shuxia. 2012. Day-ahead electricity price forecasting by a new hybrid method. *Computers & Industrial Engineering*, **63**(3), 695–701.

Ziel, Florian. 2016. Forecasting electricity spot prices using LASSO: On capturing the autoregressive intraday structure. *IEEE Transactions on Power Systems*, **31**(6), 4977–4987.

Ziel, Florian, Steinert, Rick, & Husmann, Sven. 2015. Efficient modeling and forecasting of electricity spot prices. *Energy Economics*, **47**, 98–111.