UNIVERSITY OF LJUBLJANA
FACULTY OF ECONOMICS

MASTER'S THESIS

A SHORT TERM FORECASTING OF ELECTRICITY DEMAND: THE CASE OF SLOVENIA

Ljubljana, September 2016 BORIS ČUPELJIĆ

AUTHORSHIP STATEMENT

The undersigned Boris Čupeljić, a student at the University of Ljubljana, Faculty of Economics, (hereafter: FELU), author of this written final work of studies with the title A short term forecasting of electricity demand: the case of Slovenia, prepared under supervision of prof. dr. Igor Lončarski

DECLARE

1. this written final work of studies to be based on the results of my own research;
2. the printed form of this written final work of studies to be identical to its electronic form;
3. the text of this written final work of studies to be language-edited and technically in adherence with the FELU's Technical Guidelines for Written Works, which means that I cited and / or quoted works and opinions of other authors in this written final work of studies in accordance with the FELU's Technical Guidelines for Written Works;
4. to be aware of the fact that plagiarism (in written or graphical form) is a criminal offence and can be prosecuted in accordance with the Criminal Code of the Republic of Slovenia;
5. to be aware of the consequences a proven plagiarism charge based on the this written final work could have for my status at the FELU in accordance with the relevant FELU Rules;
6. to have obtained all the necessary permits to use the data and works of other authors which are (in written or graphical form) referred to in this written final work of studies and to have clearly marked them;
7. to have acted in accordance with ethical principles during the preparation of this written final work of studies and to have, where necessary, obtained permission of the Ethics Committee;
8. my consent to use the electronic form of this written final work of studies for the detection of content similarity with other written works, using similarity detection software that is connected with the FELU Study Information System;
9. to transfer to the University of Ljubljana free of charge, non-exclusively, geographically and time-wise unlimited the right of saving this written final work of studies in the electronic form, the right of its reproduction, as well as the right of making this written final work of studies available to the public on the World Wide Web via the Repository of the University of Ljubljana;
10. my consent to publication of my personal data that are included in this written final work of studies and in this declaration, when this written final work of studies is published.

Ljubljana, ________________ Author's signature: _________________________

TABLE OF CONTENTS

LIST OF TABLES

TABLE OF FIGURES

INTRODUCTION

Electricity is a very specific commodity. Due to its nature of non-storability (currently there is not cost effective way to store electricity), what is produced it is automatically consumed either by households or by the industrial sector (or some other type of consumers). From that it is easy to conclude that electricity demand (electricity load) is one of the key factors that define the prices in various electricity markets. In today's dynamic world and market liberalization, it is crucial to have on time and good electricity demand forecasts. Transmission system operators (hereinafter: TSOs), retailers, producers, electricity traders and their short-term strategies, depend on the information what the will be the demand of electricity in the next hour, day, week, month, year.

Forecasting day ahead is classified into (very) short-term electricity demand forecasting. Research has shown that there are not many publicly available forecasts of the electricity demand. Usually only publicly available forecasts for day-ahead are published by TSOs. There are a few problems regarding those forecasts from the perspective of the day-ahead trading. Most of those forecasts are published quite late. For example, for Slovenia day ahead forecast is published after 10 a.m., for Serbia is published either in the late afternoon or on the same day, some countries do not even publish their forecasts (or they publish for previous days). Also those forecasts are being adjusted without any notice (Transelectrica – Romania, Mavir – Hungary are adjusting their forecasts after publishing first version). Methodology behind those forecasts is only known to TSOs who usually are the ones who are making those forecasts.

For trading companies this is a real problem, and that problem is noticed by many companies that are offering their forecasting products to energy traders (for example, Thomson Reuters, Meteologica, Markedskraft and others). Other option is that a trading company develops its own in-house models and use them alone or in conjunction with others. Reason is that the commercial models, which are currently offered on the market, are basically black boxes. Customers do not have information (or they get limited information) how models are working, which information is being used in the model and similar. Electricity demand modeling and forecasting (and price forecasting) are quite popular topics and one can find many articles. Weron (2014) suggests two main approaches to demand modeling and forecasting, artificial intelligence and statistical methods (other categories are mostly used in price forecasting). Similar day method or naïve method is the simplest one and most of the time used as a benchmark to other methods. Another simple method is exponential smoothing (Taylor, 2003). Regression methods (Clements, Hurn, & Li, 2016) are the most often used techniques. Time series models like AR, ARX, AR(I)MA(X), SAR(I)MA(X) and similar are also often used. It is a very known fact that weather variables influence the daily demand and they are used as exogenous variables in numerous articles (Ružić, Vučković, & Nikolić, 2003; Clements et al., 2016). In the artificial intelligence category most popular are neural networks.

In this thesis we propose a statistical model for forecasting day-ahead electricity demand. We modeled and tested our model on the most recent real data and compared it to the currently publicly available day-ahead electricity demand forecast which is used by Slovenian TSO ELES. We based the specifications of the model on the theoretical research that we conducted. Also, we simulated the practical limitations in the forecasting procedure since the aim of the thesis is that the proposed forecasting model can be used in practice and results of the forecasts should be at least equal to the currently used forecasts. Although proposed model is not significantly better than the benchmark, results showed that our model is good enough to be used in practice.

The thesis consists of the following sections. We will first present a short overview and specifics of the electricity markets and the electricity demand itself. Also, we will discuss external factors which influence short-term electricity demand. Short examples from the region about the availability of data will show the need for the in-house model building as an additional benchmark (or only for some countries), which was the basic motivation to present a potential solution to this problem.

In the data mining and modeling section, we will discuss the data mining process, also the process of identifying errors in the data, and aggregating time series for the modeling procedure. Also, in this section we will discuss the modelling procedure.

In the empirical section we will propose two models with the same data and variables, but with different approach to hourly data, and compare their forecasting results. We will compare the better model to the existing benchmark. Results of the comparison will show if our model is "good enough" or better than the benchmark.

In the final section we conclude with the summary of the results and discuss potential future developments.

1 RESEARCH AND THEORETICAL FRAMEWORK

1.1 Economics of electricity

Electricity demand (load) is defined by the behavior of the population of some country, their industry and country infrastructure. All activities of one country are considered as aggregated electricity demand (or just electricity demand) of some country. Since the electricity still does not have any cost effective way to be stored, population behavior has direct effect on the electricity demand, and in the end, on the electricity price. That also means that the produced electricity is consumed at the same time, and because of that in literature and in practice, electricity demand is also called electricity consumption or electricity demand (Do, Lin, & Molnár, 2016).

As any other price, electricity price is determined by the supply and demand. In this case, suppliers are producers or power plants, and the demand represents the population of some country, their industry and infrastructure.

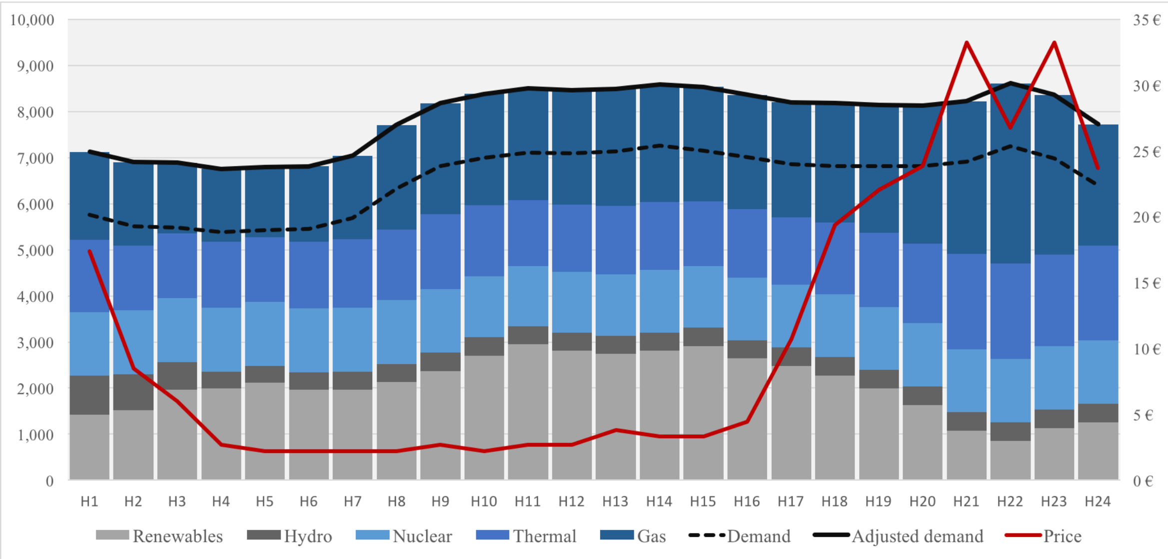
There are different types of power plants. The main types of power plants depend on the fuel which they use to produce electricity (for example nuclear, thermal or gas power plants) or which renewable source they use (for example hydro, wind or photovoltaic power plants). That usually means that their producing price is directly linked to the main source of energy. In case of some long-term contracts or some other contracts (reserves and similar), that is not the case since the price has been already defined by those special cases. Since power plants are also business entities (although a lot of power plants are state owned), they try to behave rationally and to work only when the electricity price can cover their costs.

The supply stacks in figure 1 and figure 2 are simplified examples of the intersection of the supply and demand on a day-ahead market in Romania. Both figures are showing working days, but with two different scenarios. On the horizontal axis are shown the hours of the day; the left vertical axis is the production amount in megawatt-hours (hereinafter: MWh); and the right vertical axis is the price. Figure 1 shows the production stack with very high renewables (Romania's main source of renewables is wind production) and figure 2 shows the production with very low wind and higher demand (dashed line) for approximately 500 megawatts (hereinafter: MW) in peak hours (the standard peak product is defined as the average from hour 9 to hour 20). Romania's production consists of:

- nuclear power plants,
- thermal power plants,
- gas power plants,
- hydro power plants,
- renewable production (wind, photovoltaic and biomass production).

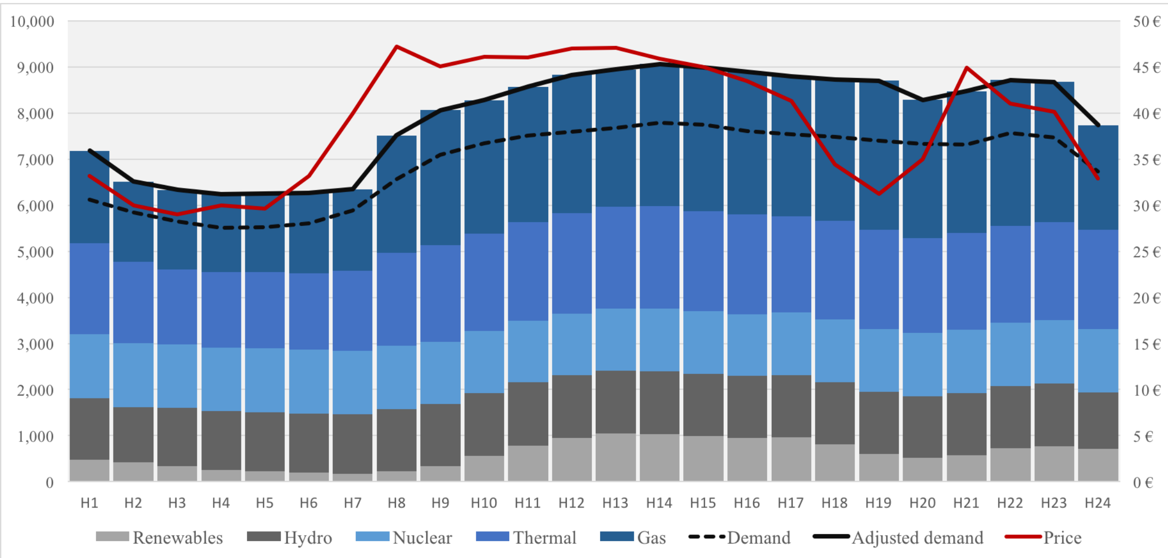
When there is a lot of renewables production, the price is much lower. Also, that day's demand is lower comparing to figure 2. Although the price order of the production might not be correct, when one compares both figures, it is noticeable that levels of thermal, hydro and gas production in the second scenario are much higher. More power plants (or the same power plants but in full capacity) had to work to cover increased demand and also to replace less wind production. Those two effects are the consequence of such a price change. What is also interesting is that those two events are less than week apart.

Figure 1. Example of day-ahead prices and realized consumption and production with high renewable production for Romania in 2016



Source: *Transelectrica. Data for electricity demand, production, exchange*, 2016; *OKTE. Total STM results CZ – SK – HU – RO*, 2016.

Figure 2. Example of day-ahead prices and realized consumption and production with low renewable production for Romania in 2016



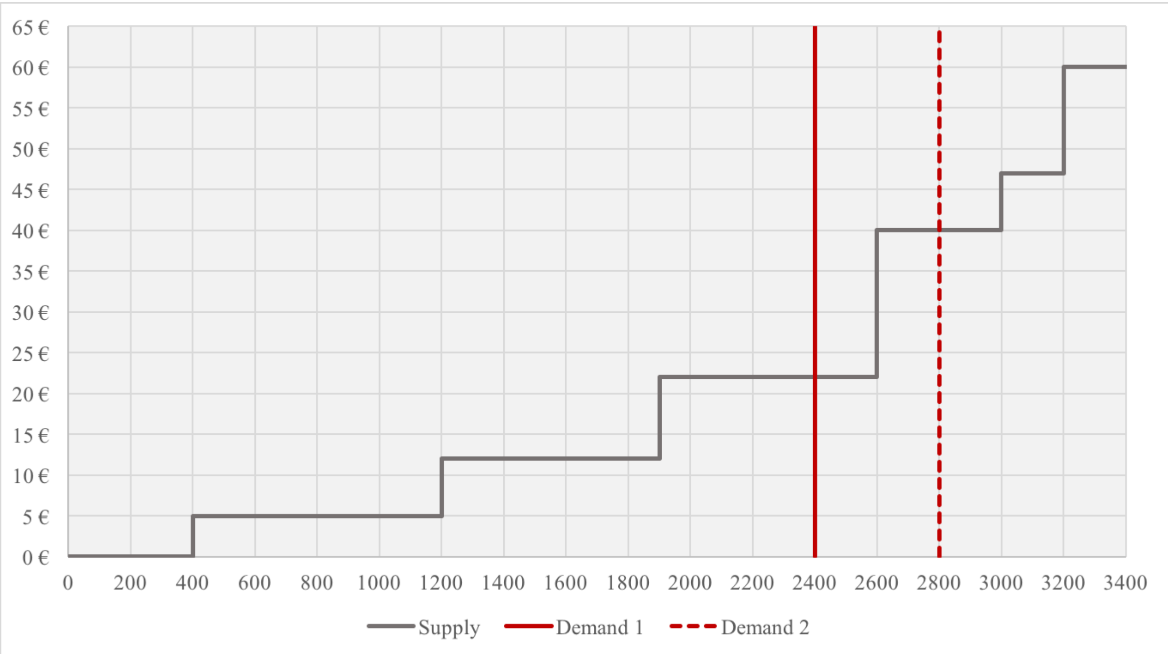
Source: *Transelectrica. Data for electricity demand, production, exchange*, 2016; *OKTE. Total STM results CZ – SK – HU – RO*, 2016.

Romanian day-ahead market price is calculated as an average from all 24 hours. That means that for each hour demand and supply are different. Figure 3 shows an example of a price formulation for a single hour. In this case vertical axis is the price and horizontal axis is production and demand in MW. Dark grey line represents a supply curve, while dark red lines represent two different electricity demand levels.

The supply curve has different levels depending on the production type. It shows ordered available production and prices for each production type. This is called a "merit-order". Although some authors (Sensfuss, Ragwitz, & Genoese, 2008) use the merit-order mostly to show how renewables production shifts the supply curve and with that influences the final price, we will assume the same production structure and different demand scenarios. Renewables production is considered to have the lowest price (0 EUR) and in our example their current production level is 400 MW. The most expensive production is usually considered to be gas production, and on the supply curve, they can be easily identified by quite high price jumps. Two demand lines represent different demand scenarios. In case of lower demand (red line), the price is 22 EUR/MWh, but if the demand is higher, the price jumps to 40 EUR/MWh.

To make a similar comparison like in the first example, we can change any of the production stack from the merit-order. If we increase any production levels left from our demand lines, it automatically shifts whole production to the right and leads to lower prices. That happens when there is a lot of renewables production, like high wind or precipitation (run-of-river production does not have any accumulation and because of that the amount of production is directly correlated with the precipitation). In the case when some certain production unit is unavailable, it shifts the merit-order to the left and increases the price.

Figure 3. Example of supply and demand curves for a single hour



From the aspect of production (at least in the short-term), electricity demand is very inelastic, it is given. Depending on the options how to "cover" demand, the final price is formed. Although these two examples are quite simplified because we have ignored the transfer of electricity between countries they show how much demand is important in electricity markets and how electricity price markets can be very volatile in short-term trading.

1.2 Specifics of the electricity demand

Before, forecasting of electricity demand was only important to electric utilities and power plants. Now, when markets are deregulated, demand forecasting is important to all market participants. That is why now electricity demand is analyzed even more in detail. Here we will present some specifics of electricity demand and external factors which have influence on short- and long-term basis.

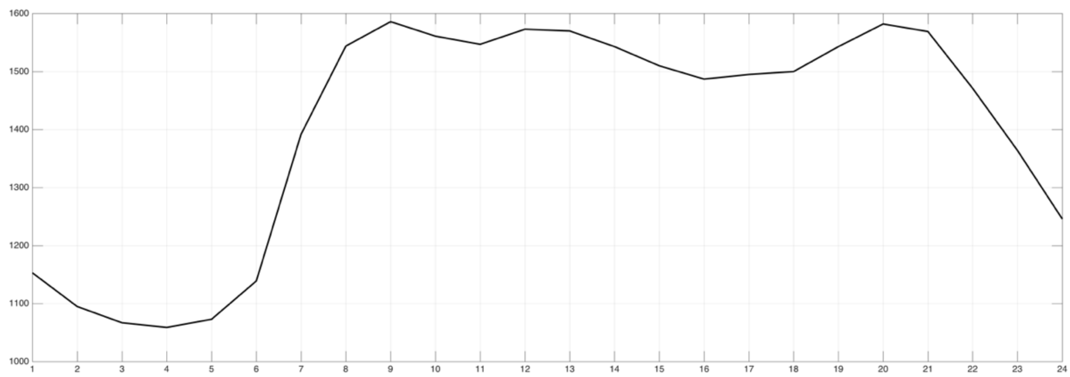
1.2.1 Seasonalities in electricity demand

Electricity demand has a specific "shape", which usually differs from country to country, depending on the size of the population, industry specifics, and some other macroeconomic variables. There are very noticeable three seasonalities, daily (hourly), weekly, and yearly (monthly). All three are important to recognize in order to properly model and forecast electricity demand.

1.2.1.1 Daily demand

Daily activities of the population can be easily recognized in the shape of the hourly demand. An example of the daily hourly demand for working days (without holidays) can be seen in figure 4. On horizontal axis are hours of the day, while on vertical axis is the consumption in MWh. During the night hours, when there is least activity, demand is lowest. People usually wake up between six and seven in the morning and prepare for work. In that time period, the slope of the demand starts to be steep. As more and more people are going to work, industrial production is starting. All those activities are forming "the first shoulder" in the daily curve. Depending on weather factors (which will discuss later), the level of middle (peak) hours varies. Around 4 p.m. people usually finish with work and go home and rest for a few hours. In that period, there is a drop in electricity demand. In the late evening most of the people focus on their household activities like cooking, cleaning, watching television and similar which form "second shoulder" in the daily curve. In the late evening, when people go to sleep and the second industrial shift ends, the demand starts to drop.

Figure 4. Example of hourly electricity demand shape for Slovenia

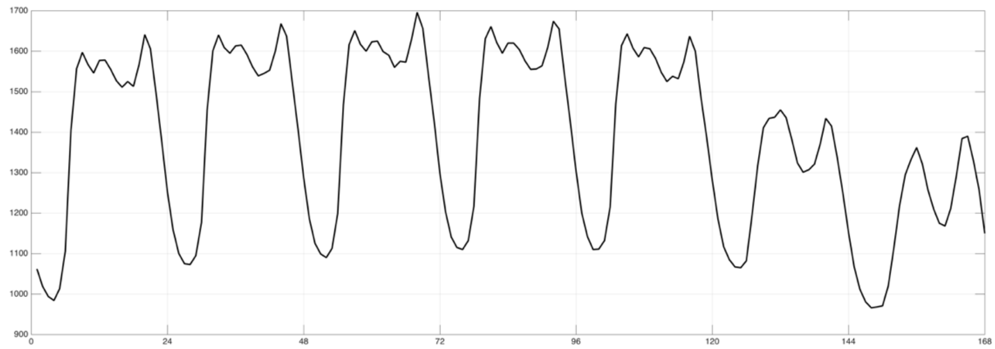


Source: *ELES. Load and Generation*, 2016.

1.2.1.2 Weekly demand

Daily demand shape is different for almost every day in the week (figure 5). Horizontal axis has the hourly scale, divided by the days in a week (from Monday to Sunday). Vertical axis is the consumption in MWh. The peak demand is similar for all working days, but Monday and Friday are a bit different. Since Monday is the first day in the week, the demand in the first hours is lower than for other working days. It is more similar to Sunday. On Friday, the second part of the day also tends to be lower than on other working days when people go home from work. Although most people are at home during the weekend, Saturdays have larger demand than Sundays.

Figure 5. Example of weekly electricity demand shape for Slovenia

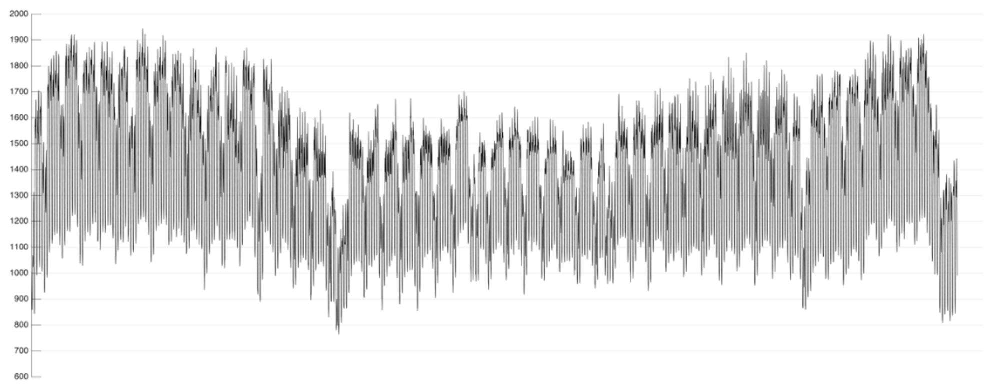


Source: *ELES. Load and Generation*, 2016.

1.2.1.3 Yearly demand

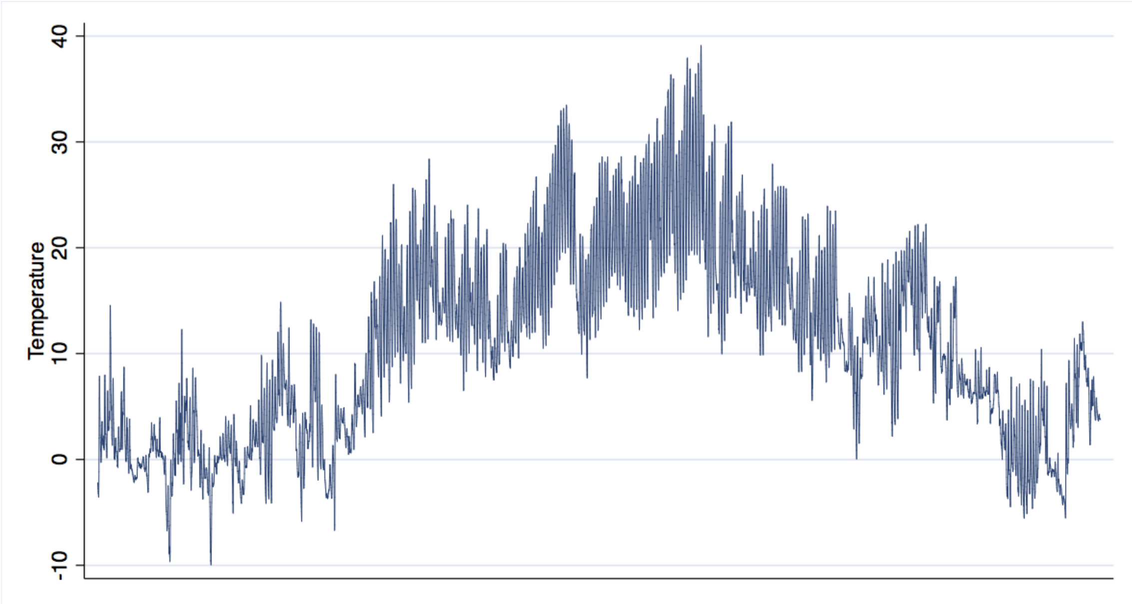
In Slovenia, yearly seasonality is not as much pronounced as in some other countries, but it is noticeable. Figure 6 shows hourly demand for Slovenia for the year 2013. Vertical axis shows demand in MWh. By visual inspection, it is noticeable that demand has seasonal fluctuations due to weather factors (primarily temperature – figure 7). In colder months, demand is higher because people use more electricity for heating than for cooling in summer months. Also, some weeks have much lower demand due to holiday effects (Christmas, Easter and similar).

Figure 6. Hourly electricity demand for Slovenia in 2013



Source: *ELES. Load and Generation*, 2016.

Figure 7. Hourly temperature for Slovenia in 2013



Source: *Ministry of the environment and spatial agency - Slovenian Environment Agency*, 2016; own calculations.

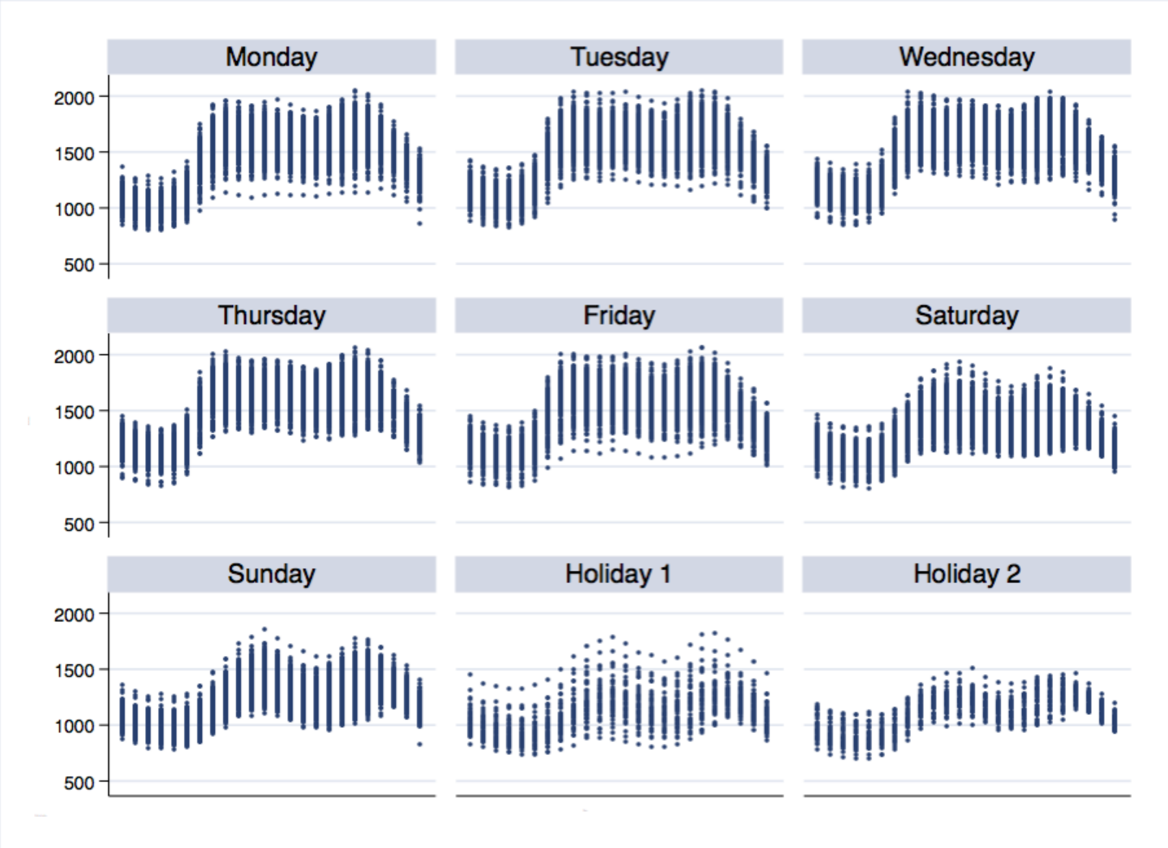
1.2.1.4 Special days

Special events can be defined as abnormal days, during which demand is quite different from its normal shape. They are also rare events and, because of that, more difficult to forecast. These days are mostly public holidays.

Not all holidays have the same effect on demand. From that perspective, holidays can be categorized as main holidays and other holidays. Also, holidays can have an effect on the next day and/or the day before. One of the examples is Serbia. If a particular holiday falls on weekend (for example a Saturday), holiday effect is prolonged on the first normal working day, but that holiday has the effect on Saturday and Monday, and also on Sunday. Other example would be if holiday falls on Thursday, demand on Friday could be lower compared to a normal Friday due to extended weekend effect. In some countries, such days are considered as holidays. They are called bridge days. Apart from normal holidays, which lower the consumption, there are special events which increase the consumption. One of those events are working Saturdays. Since bridge days are not regular non-working days, to compensate for them, Hungary has working Saturdays.

Figure 8 shows scatter plots of realized demand in Slovenia by hours, for each day in a week and also holidays as special days (holiday 1 and holiday 2). We can see that holidays have different demand levels than any other day. The shape is similar to Sunday, but there is a noticeable difference, where holidays are mostly lower than Sundays. One can also see a few lower shapes on Monday and Friday, which indicates the existence of bridge days. The proof that holidays have lower demand compared to other days in a week is reported in table 3.

Figure 8. Hourly shapes for each day in a week with holidays as separate categories



Source: *ELES. Load and Generation,* 2016; own calculations.

Other abnormal daily patterns could be some random events, like blackouts. Because blackouts and different sorts of outages are random and irregular events, they should be considered as anomalies since they produce noise to normal demand.

1.2.2 Weather factors

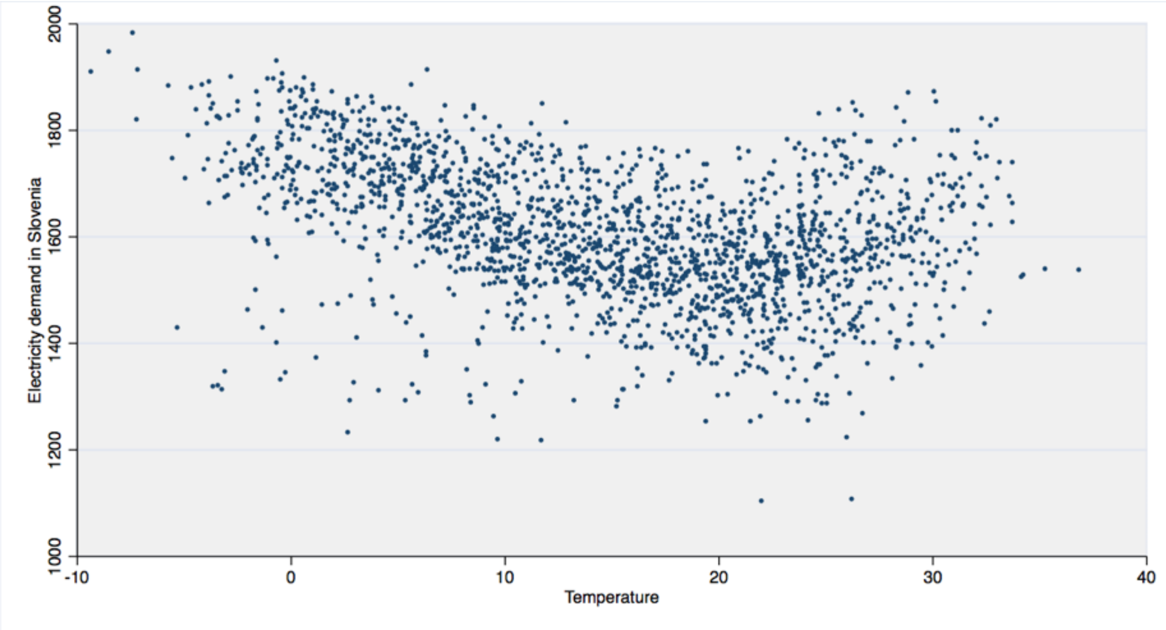
Apart from the seasonalities in the model, weather factors are very important for electricity demand forecasting. They are mostly used in (very) short-term forecasting, since long-term weather forecasts are not always available and are also quite imprecise. Temperature is the most common used weather factor. Other factors which can also be used in modeling are solar radiation, wind speed, humidity and others.

1.2.2.1 Temperature

Seasonal fluctuations during the year are due to weather factors. One of the most important weather factors is temperature. A survey by Hippert, Pedreira, & Souza (2001) on models for forecasting electricity demand shows that out of 23 papers, in 13 papers authors used only temperature, in 3 papers they used temperature and humidity, other 3 used also additional weather parameters, and last 3 did not use any weather parameters. One of the reasons why authors mostly used temperature is because many of them did not have access to other data.

Weron (2006) argues that the relationship between demand and temperature is not linear. Figure 9 shows scatter plot between electricity demand (vertical axis in MWh) and temperature (horizontal axis in Celsius). To make it more clear, it is filtered only for working days (without holidays) and for hour 13. It is clearly a non-linear relationship. He explains that the shape looks like a hockey stick. The breaking point is approximately around 19 degrees Celsius. He also suggests that quadratic form is suitable for mimicking the non-linearity.

Figure 9. Scatter plot between electricity demand in Slovenia for working days (without holidays) and hour 13 and temperature in Celsius.

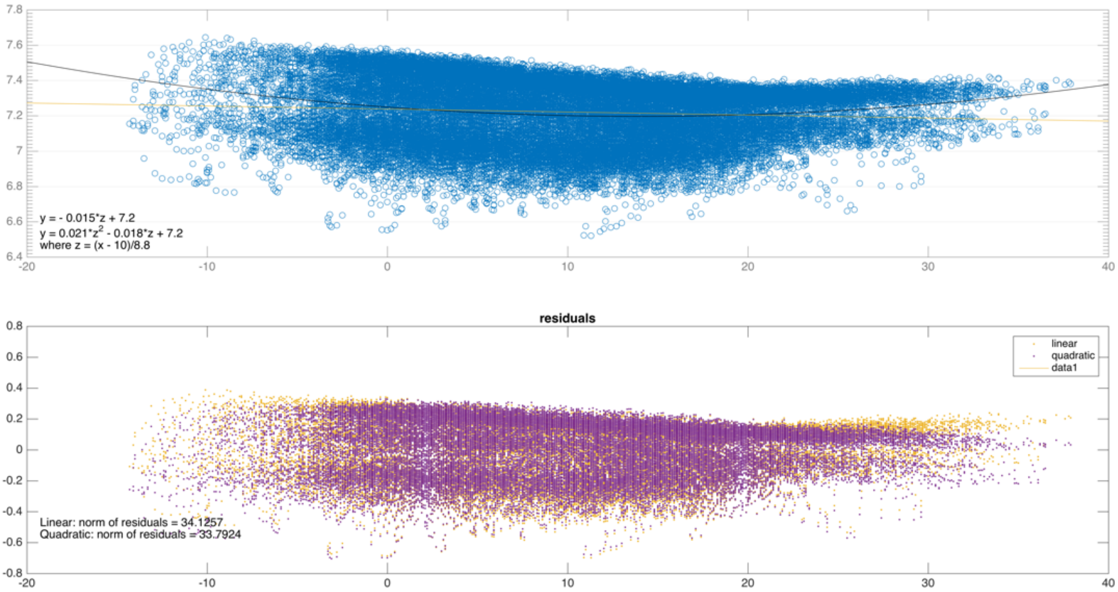


Source: *ELES. Load and Generation*, 2016; *Ministry of the environment and spatial agency - Slovenian Environment Agency*, 2016; own calculations.

MATLAB offers a simple fitting tool which helps to easily approximate which polynomial form should be used. Figure 10 shows an example of the MATLAB tool. It consists from the two scatter plots. Both of the scatter plot have the same horizontal and vertical axes, where horizontal axis is the temperature in Celsius and vertical axis is the demand in MWh. First scatter plot shows different fit profiles, in this case linear and quadratic fit. Second scatter plot shows the residuals of each fit profile. One can also check polynomials of higher degree in this tool. It is visible that quadratic form has better fit than linear where residuals are closed to zero in higher temperatures.

Other solution is to use a piecewise linear regression. Some authors suggest using heating and cooling degree days (Valor, Meneu, & Caselles, 2001; Pardo, Meneu, & Valor, 2002; Hor, Watson, & Majithia, 2005).

Figure 10. MATLAB curve fitting tool example



Source: *ELES. Load and Generation*, 2016; *Ministry of the environment and spatial agency - Slovenian Environment Agency*, 2016; own calculations.

1.2.2.2 Solar radiation

Solar radiation, in plain words, shows how much light is reflected on Earth's surface. In Slovenia, Slovenian Environment Agency (hereinafter: ARSO) measures two type of solar radiation, global (which is total radiation) and diffused solar radiation. Global radiation is a sum of direct and diffused radiation. In cloudy weather, when there is a little or no direct solar radiation, diffused radiation is important. Levels of global solar radiation are naturally dependent on the period in the year.

As any other weather variable, there is no local measurement for the whole country. It is measured by number of weather stations throughout the country. In the beginning of 2008, ARSO had 29 out of 45 stations which measured solar radiation.

One of the replacements for solar radiation the electricity trading participants use are photovoltaic (solar) production forecasts. They measure the levels of solar energy production for the whole country. Since solar radiation and photovoltaic production are directly correlated, forecasts of photovoltaic production are based on forecasts of solar radiation. Another reason why traders use derived forecast is to be able to forecast the residual load. Residual load is defined as the demand from which is subtracted renewables production (like photovoltaic and wind). In countries with very high renewables production (Germany), the analysis of residual load is very important in price forecasting.

1.2.2.3 Other weather factors which influence the electricity demand

There are two weather factors which some authors additionally use (Hyde & Hodnett, 1997; Fay, Ringwood, Condon, & Kelly, 2003), humidity and wind. Humidity and wind are mostly used to enhance the effect of the main weather driver, the temperature. They could also help to explain non-linearity between the temperature and the electricity demand. During summer days, when temperatures are high, high humidity increases temperature effect. In consequence, people use more and air-conditioning longer to cool. In winter days, when it is cold, high wind speeds "help" to cool buildings even more, so people use additional electricity to get warmer.

1.3 Overview of forecasting models

We can classify forecasting models from a few perspectives. One perspective is how far ahead one forecasts and they are classified as:

- (very) short term forecasts,
- medium term forecasts,
- long term forecasts.

Very short term forecasts to short term forecasts have a time span from 15 minutes (for example for continuous intraday 15 minute markets in Germany or France - EPEX SPOT) up to one week ahead. With large amount of renewable energy like wind or photovoltaic (solar power production), hourly products are not enough. Intraday markets like 15 minutes ahead are developed for the balancing purposes (in case some market participant have excess or require energy due to changes in electricity demand or in production). Day ahead demand forecasts up to one week are usually used for creating strategies and price forecasts for short-term trading. In short term forecasting, weather variables have a great impact on the electricity demand.

From the trading perspective, time periods from one week to one year are considered as medium term forecasts. For shorter time periods like week up to month ahead it is possible to get weather forecasts, but for longer time periods only seasonal effects can be modeled.

For long term forecasts it is practically impossible to find precise weather forecasts. However, it is possible to provide the probability distribution of the electricity demand based on historical weather observations (Feinberg & Genethliou, 2005). Large historical datasets (for example past 20-30 years of data) are used to create "normal weather variables" (in simplistic terms average historical weather) to forecast weather normalized demand. Depending on the aim and length of long term forecasts, one can also use macroeconomic variables (industry growth).

Other perspective to electricity demand forecasting is the method which is used. There are many different forecasting models, but they can be easily classified into two main categories, artificial intelligence models and statistical models. One of the classifications is presented in Alfares and Nazeeruddin (2002). Weron (2014) presents a nice classification of models for electricity price forecasting. From that paper last two categories are also used in electricity demand forecasting. Suganthi and Samuel (2011) have similar categorization.

1.3.1 Artificial intelligence models

In the artificial intelligence category, the most popular models are artificial neural networks, support vector machines, fuzzy logic and similar. Artificial intelligence models are known for their ability to cope with complexity and non-linearity. Hippert et al. (2001, p. 46) define artificial neural networks as: “Artificial neural networks are mathematical tools originally inspired by the way the human brain processes information.”. Data is divided into three sections: learning (training), validation and estimation. Coefficients are automatically estimated through the algorithms. That is also a possible main issue for artificial intelligence models. Since learning is being done with hidden layers connecting neurons, the whole process is a black box. Also there is a possibility of overtraining the model in order to get high r-squared and bad out of sample forecasts.

Disregarding possible downsides of artificial intelligence models, they are quite popular in forecasting literature. Mostly they are used by people with technical background, where economists are more prone to statistical models. Since the main focus for this thesis is in the field of statistical models, deeper analysis of artificial intelligence models is not necessary.

Chen and Chang (2004) propose a winning model for the European Network on Intelligent Technologies for Smart Adaptive Systems (hereinafter: EUNITE) competition in 2001 for medium-term forecasting using a support vector machine. Their task was to forecast Slovakian daily maximum for a month ahead (predicting a maximum daily demand for next 31 days). They also discuss usage of weather factor in medium-term forecasting and they do not find any evidence that weather factors could improve their model’s accuracy.

Some authors conclude that artificial intelligence models should not have any advantages to statistical models. Darbellay and Slama (2000) test whether neural networks models are indeed better in forecasting when compared to statistical models, in their case ARMA type. They test their models on Czech electricity demand and they find that most of the electricity demand problems are linear in nature. Because of the linearity, there should not be an advantage in using neural networks, which handle non-linearity better than statistical models. Linear models in their case perform better at forecasting hourly demand than the proposed feed-forward and recurrent artificial neural network.

1.3.2 Statistical models

Statistical models are a mathematical model which shows electricity demand as functions of various factors. By type, there are two categories, additive and multiplicative models. Difference between those models is that additive models are a sum of various factors, and multiplicative models are a product. Although multiplicative models can be transformed into additive models (by using logarithms), additive models are much more popular. A group of statistical models includes: simple one like similar-day method, exponential smoothing, linear regression, autoregressive models and others.

1.3.2.1 Similar-day

Similar-day (also called "naïve method") approach is the easiest method to implement. It is based on searching for similar days in history (type of the day, temperature, potential holiday effect, etc.) and using them as forecasts for either one of the results or some kind of average. Electricity demand of Tuesday or of previous Wednesday can be used as the forecast for Wednesday (since both days are working days). Same logic is applied on weekends. Last Saturday or Sunday can be used as the forecast for the next weekend. Naïve models are useful as an additional tool in forecasting rare events like holidays. Sample size of holidays is usually very small, especially if available historical data is not long, advanced models could have difficulties with estimating holiday effects. Due to its simplicity it can be used as one of the benchmarks for more complicated models but it is also not recommended since naïve models tend to be imprecise.

1.3.2.2 Exponential smoothing

Exponential smoothing uses exponentially weighted averages of past observations to forecast future values. Due to its robustness and accuracy, exponential smoothing is quite often used as a forecasting method in various fields.

In simple exponential smoothing, exponentially smaller weights are assigned to older lagged values. Basic formula is

$$\widehat{L_t} = \alpha L_{t-1} + \alpha(1-\alpha)L_{t-2} + \alpha(1-\alpha)^2 L_{t-3} + \cdots \tag{1}$$

where $\widehat{L_t}$ is forecasted demand at time t, L_{t-n} are historical realizations and α is a smoothing parameter.

Because simple exponential smoothing cannot capture trend nor seasonalities, it is not a good forecasting method to forecast electricity demand. To accommodate for those trends and seasonalities more advanced models have been developed (Holt, 1957; Winters, 1960; Chatfield & Yar, 1988; Taylor, 2003).

In his paper Taylor (2003) presents adapted double seasonal exponential smoothing forecasting model on the sample of half hourly-demand for England and Wales in 2000. Because short-term electricity demand has two seasonalities, daily and weekly, standard exponential smoothing is not enough since it cannot capture two seasonalities. Taylor adapts the Holt-Winters exponential smoothing that it includes seasonalities in the model. He compares the model with the standard Holt-Winters and also to well-specified double seasonal ARIMA model. Adapted double seasonal exponential smoothing model outperforms the other two models.

In a similar study Taylor, De Menezes, & McSharry (2006) test very short term forecast models up to a day ahead. Apart from the adapted double seasonal exponential smoothing model from the paper in 2003, they also include a new model which is based on principal component analysis. Other models which they consider are naïve benchmark, naïve benchmark with an error model, a neural network model and a seasonal ARIMA model. They use two samples for model comparison, hourly time series of Rio de Janeiro and half-hourly demand for England and Wales. Although principal components analysis model performs well, the best results are again achieved with the double seasonal exponential smoothing model.

1.3.2.3 Multiple regression

Regression is one the most commonly used in electricity demand forecasting. Main purpose of the regression is to learn about the relationship between the dependent variable (in this case electricity demand) and independent variables (like weather factors, type of the day, hour of the day and similar). Estimation of the coefficients is done by using least squares.

$$L_t = \alpha + \beta_0 \cdot X_t^1 + \beta_1 \cdot X_t^2 + \cdots + \beta_n \cdot X_t^n + \varepsilon_t \tag{2}$$

In equation (2) α is a constant, β_n are vectors of coefficients, while X_t^n are exogenous variables (matrices) and ε_t is an error term. One of the assumption of the multiple linear regression is the linearity between the variables. That is why it is important to check the scatter plot of the variables (figure 9). In case it is evident that the relationship between variables is not linear, one could try to transform one of the variables.

Engle, Chowdhury, & Rice (1992) propose a model for forecasting the hourly peak electricity demand for one day in the future. Model includes deterministic variables such as holidays and lagged holidays variables, stochastic variables such as past daily average demands by building bivariate models, and also weather variables which is given a careful they transform to account for non-linearity. They tested the model on the one year out-of-sample period.

Hyde and Hodnett (1997) propose a multiple regression model for the Irish electricity supply system. Proposed model is built for day ahead forecasting and it can be also used for seven to ten days ahead. It identifies a normal or a weather-insensitive demand component and a weather-sensitive component. For the estimation of the normal demand model, they use linear regression of past demand and weather data.

Ramanathan, Engle, Granger, Vahid-Araghi, & Brace (1997) propose a short-run forecasting model of hourly system load and evaluate the forecast performance. The model is applied to historical data for the Puget Sound Power and Light Company who did a comparative evaluation of various forecasting models for two years in a row. Their approach is based on a multiple regression model. Each hour of the day for workdays and weekends is modeled separately, which results in 48 models. Simple model structure consists of four types of variables, deterministic variables such as day in a week, month and similar, weather variables such as temperature and historical temperature, historical realized dependent variables (electricity demand) and historical forecast errors (by using a fifth-order Cochrane-Orcutt autoregressive error structure). In the first step, they estimate the model by ordinary least squares (hereinafter: OLS). Then, they predict the residuals. In next step, they introduce lagged residuals into the model, and they estimate again with OLS. Models performed extremely well comparing to models of other authors. In the second year, authors of other models had the possibility to revise their models, and even then, their models were not still able to have better performance compared to authors' models.

Ružić et al. (2003) propose a regression-based adaptive weather sensitive short-term load-forecasting algorithm, which was developed and used in Electric Power Utility of Serbia (EPS). The proposed methodology consists of two main steps. In the first step, authors forecast the total daily demand independently, then, in the second step they predict hourly demand. All model parameters are automatically calculated and updated using realized data in the identification period.

Fan and Hyndman (2012) propose a semi-parametric additive model to estimate the relationships between the demand and the driver variables. Input for these models are calendar variables (day of the week, holiday effect, day in a year), lagged actual demand (demand for the same hour for past two days, maximum and minimum demand for previous day as well as average demand for previous seven days) and temperature (forecast of the temperature, minimum and maximum from the previous days, lagged half-hourly temperature as well as the average temperature for last seven days) from one or more measuring points. The proposed methodology has been used to forecast the half-hourly electricity demand for up to seven days ahead for power systems in the Australian National Electricity Market. They validate the performance of the methodology with out-of-sample forecasts with real data from the power system, as well as through on-site implementation by the system operator.

Do, Lin, & Molnár (2016) propose a regression based model on the case of Germany. They want to see whether forecasting models are more precise if electricity demand is modeled for each hour independently or all hours are modeled together. In the second model they forecast electricity demand in two steps. In the first step authors forecast average daily consumption and in the second step for each hour they forecast deviation from the average. Both models have as similar variable specification as possible to be able to compare them. Explanatory variables which they use are temperature, industrial production index, hours of daylight, binary variables for days in a week, binary variables for months of the year. They separate holidays into two categories: major (official holidays, non-working days) and minor (local or religious holidays). For holidays they use separate binary variables and also bridge days as independent variables. Industrial production index helps to model the yearly trend. With a bonus of simplicity for having less variables in a regression, independent separate hour models show better results compared to more complex model.

1.3.2.4 Autoregressive model

Assumption of time series models is that they have an internal structure, like trend, seasonality and autocorrelations. Autoregressive models are designed to exploit those features and use them to forecast future values. An AR(p) model indicates an autoregressive model of order p.

$$L_t = c + \sum_{i=1}^{p} \phi_i L_{t-i} + \varepsilon_t \tag{3}$$

where L_t is electricity demand at time t, c is an constant, ϕ_i are coefficients of the model and ε_t is a white noise. The order of p is telling us how many lagged past values are included into a model.

1.3.2.5 Autoregressive (Integrated) Moving Average model

The moving average is a model where the time series is regarded as a moving average of previous error terms ε_t. A moving average MA(q) model of order q is given by

$$L_t = \mu + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{4}$$

where μ is a constant and other terms are known from previous equation.

ARMA processes require that the time series which is being modeled is (weakly) stationary. Stationarity is defined that first and second moment should not change over time. If it is not, then one should transform the data to make it stationary (for example by using first difference or more).

Box and Jenkins (1976) introduce a general model which contains both AR and MA parts and also a differenced part (Integrated – I, usual notation for order is *d*). Sometimes, difference is not enough due to seasonality so additional factor was needed. Seasonal Autoregressive Integrated Moving Average model (SARIMA) is introduced. General notation is

$$\mathrm{SARIMA} = \mathrm{ARIMA(p, d, q)}\times\mathrm{(P, D, Q)_s}. \tag{5}$$

1.3.2.6 Autoregressive models with exogenous variables (ARX models)

Autoregressive models are using historical signals to forecast future values. In some cases, apart from the signal itself, future values are influenced by some exogenous factors (in the electricity demand case one factor could be temperature, for price forecasts, an additional factor would be wind production, or transfer capacities and similar).

In his article Weron (2014) also discusses difficulties with categorization or differentiation between regression and ARX models (article is focused on overview of forecasting models for electricity prices, but same logic can be applied to forecast the electricity demand). He says that it is often very hard to separate regression and autoregression models since a lot of models are called regression models, but they contain previous values of the dependent variable. He suggests that difference between those two models could be made from the structure of the model. If the number of the fundamentals regressors is large, then it should be classified into regression models, and if the autoregressive structure is complex, then they should be classified as autoregressive models with exogenous variables.

2 DATA AND METHODOLOGY

2.1 Demand data

Slovenian electricity transmission system operator – ELES publishes two types of data for electricity demand in Slovenia, electricity demand per type and aggregated demand data. There are two types of consumers, direct consumers and consumers via distribution. Second category is more than 80% of the total demand. Aggregated demand data also includes losses on the transmission network. Consumption of pump storages is included in the production data which makes data mining and forecasting procedures easier.

Realized hourly data is published with a few hours of delay (most of the time not more than two hours of lag). Every working day after 10 a.m. ELES publishes its own hourly forecast for day-ahead demand (on Friday late afternoon forecasts for Sunday and Monday are also available). On their website there is no available information about the forecasting model. Although electricity market for day-ahead in Slovenia is open until 12 p.m., that time for trading companies is quite late because all forecasts and strategies for day-ahead are usually made much earlier. Data for day-ahead and analysis larger period can be exported to Excel (it is HTML table data with a MS Excel extension), XML and CSV formats.

2.1.1 Some other countries

Each country has a different methodology of making and publishing their electricity demand forecasts. In wholesale trading those differences make those forecasts mostly unusable or quite risky to use. First and main issue is that some countries publish their forecasts quite late for day-ahead trading. Second issue is that methodology behind those forecast is unknown, which parameters are taken into consideration and similar. Third issue is that TSOs are changing their forecasts without any notice which leads to losing track of the precision of their models. One more issue which must not be ignored is availability of the data. Sometimes forecasts are either not published or websites are not available. Below are examples for few countries.

2.1.1.1 Serbia

Data for Serbian demand can be downloaded from EMS[1] Transparency website. Other source is Entso-e. Until February 2016, EMS was publishing realized consumption without including Kosovo. Since this change was a silent update (without any notification), a lot of data and forecasts providers were not aware of it. Their forecasting models for demand were quite imprecise for a week or more since including Kosovo into total Serbian demand increased the average daily consumption for approximately 600 MW. Realization of the demand is published around 7:30 a.m. for the day before, and forecasts for day-ahead are published quite late (either after 4 p.m. or on the same day). It is also worth mentioning that if one compares EMS demand forecasts and realizations, it seems that pump consumption is only included into forecasts.

2.1.1.2 Croatia

Data for Croatian demand can be downloaded from Entso-e. Realizations of the demand is published within a few hours of delay. In case of a missing few hours on a certain day, most of the time data is being updated backward (errors in published data for yesterday are updated today). Forecasts of the electricity demand for day-ahead are available in the late afternoon.

2.1.1.3 Hungary

MAVIR publishes data on the demand and their forecasts. Frequency of the data is 15 minutes and it has possibility to be converted into the hourly demand (simple average for a specific hour). Forecasts are published around 4 p.m. for the day after tomorrow and for day-ahead previous forecast is updated. On Friday afternoon MAVIR publishes forecasts for the Sunday-Tuesday and it also updates those forecasts. Since on their website one can download only last version of the forecasts (which are published after all markets for day-ahead trading are closed), it is very difficult to measure the precision of their forecasting model of the electricity demand. Only solution how to measure accuracy of their initial forecasts is to download their day-ahead forecasts on a daily basis which is very time consuming.

2.1.1.4 Romania

Transelectrica[2] publishes week ahead hourly forecasts. Methodology of the forecasts is unknown, but by visual inspection of the forecast data, most likely naïve method is used (for example similar day from previous week). Those forecasts are updated on a daily basis at an unknown time period and also day zero (today) forecasts are updated.

---

[1] EMS stands for Elektromreža Srbije (hereinafter: EMS)

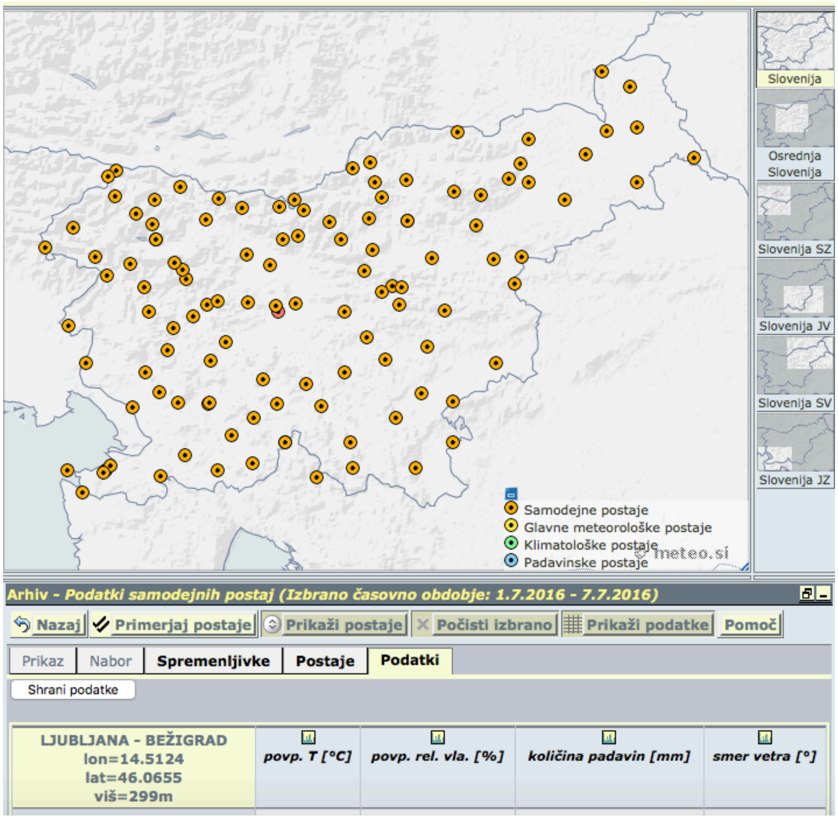
[2] One can find electricity demand forecast on *Transelectrica. Required Transparency of System Load*.

2.2 Weather data

In Slovenia, weather data is published by ARSO. Each year, number of measurement stations is increased. Depending on the type of the stations, one can find data on temperature, solar radiation, rainfall, wind speed, humidity, air pressure and others.

The most frequent data which one can download from ARSO is the half-hourly data. That data has to be converted to the hourly so that it can be used with the hourly load. Since ARSO does not have singular data for Slovenia, data from various stations also have to be aggregated into one time-series per type. Number of measuring weather stations increased during the past few years (figure 11), but we can only use stations which were in existence at the beginning of our modelling procedure.

Figure 11. ARSO weather data export user interface



Source: *Ministry of the environment and spatial agency - Slovenian Environment Agency*, 2016.

2.3 Data mining procedure

2.3.1 Dealing with erroneous demand data

To be able to make a good forecasting model or any other analysis, we have to identify and correct various data errors. The higher is the data frequency, the higher chance it has some errors. There are two type of errors, outliers (strange, most likely incorrect values, spikes) and missing data. Single missing entries should not make a problem, but missing days or weeks can cause a problem. In hourly demand data one error is quite common, such as the change from normal to the daylight savings time. Methods of reporting demand for that hour (or hours) can vary from year to year and it should be checked manually. Reasons for other outliers or missing values are unknown, but most likely they are caused by some outages and it would take time to identify them manually.

First step with dealing with outliers is to identify them. One of the methods which Weron (2006) suggests is to create an automatic filter using running median. Compared to moving average, running median is more robust to outliers. He also advises to use short- and long-term running medians, since short-term running median can only detect large spikes for single hour, and long-term should be able to detect outliers for larger periods (days). Berk (2015) is also using same filtering technique, but he suggests that each demand time series is unique and one should test different lengths of short- and long-term filters.

The (2m+1)-hour running median is given by

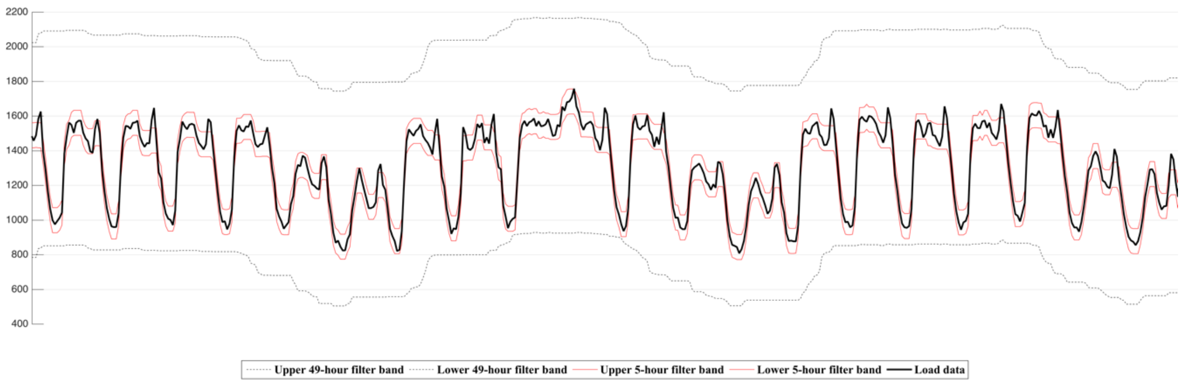
$$L_t^{\mathrm{med},2m+1} = \mathrm{median}(L_{t-m}, \dots, L_t, \dots L_{t+m}). \tag{6}$$

Next step is to create upper and lower filter bands.

$$B_t = L_t^{\mathrm{med},2m+1} \pm k \cdot \mathrm{std}\left(L_t - L_t^{\mathrm{med},2m+1}\right). \tag{7}$$

Value of k in equation above is usually 3 and std represents standard deviation. Examples of 5-hour and 49-hour filtering bands for Slovenia are shown in figure 12.

Figure 12. Example of 5-hour and 49-hour running median filtering bands for electricity demand in Slovenia



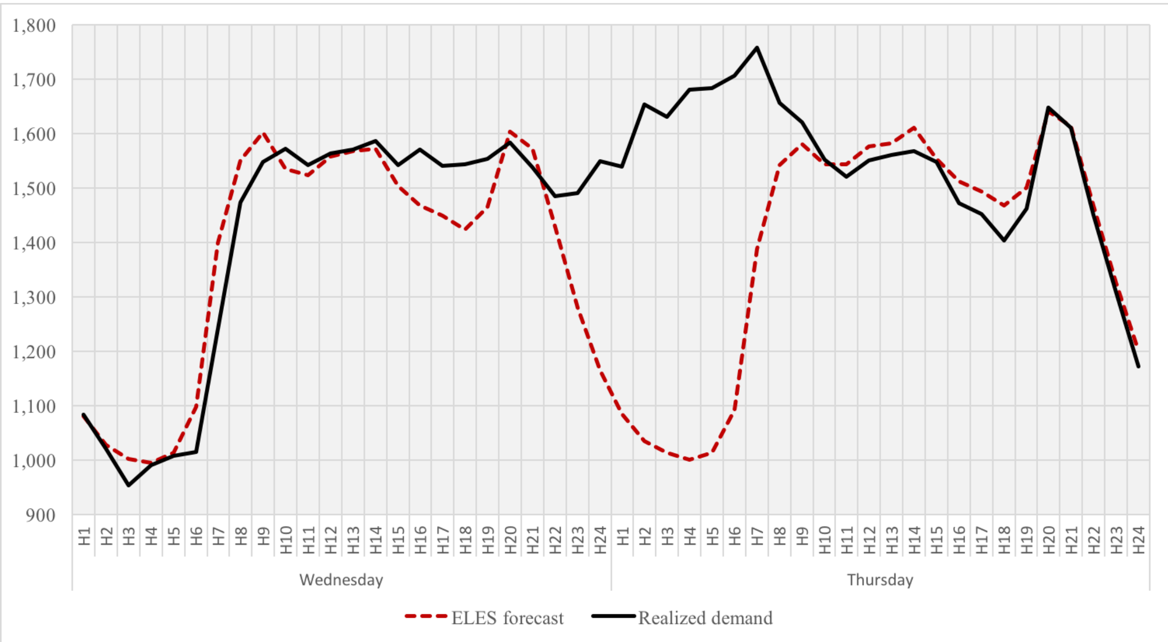
Source: *ELES. Load and Generation*, 2016; own calculations.

Once errors in data have been identified, they have to be fixed. Easiest way to deal with single missing value for period t is to take an average of the t-*1* and t+*1* observations. To deal with more missing values Weron (2006) suggests forecasting them, but since advanced forecasting methods usually cannot be estimated and calibrated to the data with missing values, one should use similar-day method instead. Forecasted values should be treated as the original ones.

Filtering models are very helpful in detecting outliers, unfortunately there is also a possibility of wrong positives. Neither short- or long-term filter detected this kind of error. In the middle of figure 12, two days are connected into one, demand in night hours is unusually high. One possible explanation is that in those hours pump production was included by mistake (which is, in case of Slovenia, included in the production data). Short- and long-term filters are not indicating incorrect data. Easiest way to replace this data is to take an average of same hours from previous and leading days (depending on the day in a week, it is also possible to choose days from previous and leading week). Since historical day-ahead demand forecasts are already publicly available for Slovenia (figure 13), they can be used as a benchmark for correcting that error.

Visual inspection of whole hourly data would be very inefficient and time consuming, therefore the correct approach is to look for sudden large deviations from the forecasts. Larger errors are more often found during the holidays, so holiday indicator should be included into the filtering procedure. In figure 13 dashed red line represents ELES's forecast of the demand, and black line is the realization of the demand. If we also inspect previous and leading days, we can easily see that those night hours are most likely incorrect data and they should be replaced by forecasts.

Figure 13. Example of incorrect data for realized electricity demand



Source: *ELES. Load and Generation*, 2016.

2.3.2 Data mining of weather data

Historical weather data is publicly available on ARSO website. Half-hourly data is the closest available timeframe to hourly electricity demand. Primary weather variables which are downloaded are half-hourly averages of temperature and global solar radiation[3]. Per each request, one can download up to three to four months of data for one measurement station.

We checked each dataset for errors. We use similar procedures to those for electricity demand. Small sets of missing data (up to a few hours) is forecasted by using a naïve method. In case of more missing data than a few hours, there are two solutions:

- first solution is to find the nearest similar measurement station and replace missing data with theirs,
- in case it is not possible to do that, ignore it in the aggregation process.

Since correlation between measurement points is quite high, Fan and Hyndman (2012) suggest to use simple average on hourly data (table 1). Other method which one can use are population weighted averages (Pardo et al., 2002).

Table 1. Correlation matrix of temperatures across Slovenia from 2010 to April 2016

	Lešče	**Bilje**	**Novo Mesto**	**Ljubljana**	**Maribor**	**Celje**	**Murska Sobota**
Lešče	1.00						
Bilje	0.94	1.00					
Novo Mesto	0.97	0.94	1.00				
Ljubljana	0.97	0.95	0.98	1.00			
Maribor	0.97	0.93	0.97	0.97	1.00		
Celje	0.96	0.93	0.98	0.98	0.98	1.00	
Murska Sobota	0.96	0.92	0.97	0.97	0.99	0.97	1.00

Source: *Ministry of the environment and spatial agency - Slovenian Environment Agency*, 2016; own calculations.

Solar radiation was collected from same data points and the procedure is the same. One can observe lower correlation for solar radiation than for temperatures. Reason for that can be quite simple. Because solar radiation measures the amount of sunlight which reaches the surface, amount of total radiation can be lower due to the cloud cover in certain regions.

[3] Definition of each weather variable (in Slovene) is available at http://meteo.arso.gov.si/uploads/meteo/help/sl/razlaga_spremenljivk.html

Table 2. Correlation matrix of solar radiation across Slovenia from 2010 to April 2016

	Lešče	Bilje	Novo Mesto	Ljubljana	Maribor	Celje	Murska Sobota
Lešče	1.00						
Bilje	0.91	1.00					
Novo Mesto	0.88	0.86	1.00				
Ljubljana	0.90	0.88	0.91	1.00			
Maribor	0.89	0.86	0.91	0.89	1.00		
Celje	0.90	0.86	0.93	0.91	0.95	1.00	
Murska Sobota	0.86	0.83	0.90	0.87	0.94	0.91	1.00

Source: *Ministry of the environment and spatial agency - Slovenian Environment Agency*, 2016; own calculations.

2.3.3 Daylight savings time

Before merging the datasets, we have to align (synchronize) time series. Daylight savings time is one of the issues.

For the electricity demand, issues are different. For each year we have to check the data manually. In March, demand for whole day could be shifted and last hour could be missing. In October, there is no 25th hour, but demand forecast can contain very high value for one specific hour. For the consistency of hourly time-series, in March we forecast shifted (missing) hour, and in October we remove double hour.

Weather data is published in Central European Time, so we shift summer hours and we also forecast in March shifted hour and, in October, we also remove double hour. This is especially important because of the number of solar hours in a year, where electricity demand in winter is increased due to less light during the day.

2.3.4 Merging the datasets

Both datasets are merged based on the time variable. Merging was done by using one to one merge in Stata with key variables: year, month, day and hour.

2.4 Error measurements

Popular error measurements which are usually used in comparing of forecasting models are:

- mean error – ME,
- mean absolute error – MAE,
- mean squared error – MSE,
- mean absolute percentage error – MAPE,
- Theil's U-statistic.

2.4.1 Mean error

Mean error is calculated as the average of the sum of all differences between realized values and forecasts. According to the assumptions of linear regression, mean should be equal to zero. Reason for that is because errors which are used in calculation are not in their absolute values, so they cancel each other. We will use mean to look at the seasonalities and try to see and understand if some periods are too underestimated or overestimated. In the formula below, $L_i - \hat{L}_i$ is difference between real values and forecasts (hat).

$$\mathrm{ME} = \frac{\sum_{i=1}^{n}\left(L_i - \hat{L}_i\right)}{n} \tag{8}$$

2.4.2 Mean absolute error

Mean absolute error is similar to mean error, but we sum the absolute differences. This measure shows us how close in absolute terms the forecasts are from real values. Because mean absolute errors is showing errors in a same scale of data which is measured, it cannot be used to compare errors of different data types. Since most of the errors are canceled in ME, this measure is more important. Other name for mean absolute error is mean absolute deviation. It is calculated as

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n}\left|L_i - \hat{L}_i\right|}{n}. \tag{9}$$

2.4.3 Mean squared error

Mean squared error is next step of MAE. It is an average of squared errors. Compared to MAE it "punishes" extreme errors because of squared errors.

$$\mathrm{MSE} = \frac{\sum_{i=1}^{n}\left(L_i - \hat{L}_i\right)^2}{n} \tag{10}$$

2.4.4 Mean absolute percentage error

Mean absolute percentage errors is used very often in measuring prediction accuracy of a forecasting method. It expresses accuracy as a percentage (if it is multiplied by 100).

$$\mathrm{MAPE} = \frac{1}{\mathrm{n}}\sum_{i=1}^{n}\left|\frac{L_i - \hat{L}_i}{L_i}\right| \tag{11}$$

Although MAPE is very popular measure, it has a few issues. It cannot be used when the realized values are zero. That issue is not applicable to electricity demand. Some papers with analysis of MAPE are done by Tofallis (2015) and by Hyndman and Koehler (2006).

2.4.5 Theil's U-statistic

One additional measure for forecast accuracy is Theil's U-statistic. It is one of the simplest measurements. It is most commonly used to compare forecasting model with the simplest forecast procedure, the naïve forecast. Result is a square root ratio between sum of squared errors of our forecasting model and the sum of squared errors from the naïve model. Any ratio which is close or greater than one should be disregarded, since it is similar (or worse) to the naïve model.

$$U = \sqrt{\frac{\sum_{i=1}^{n-1}\left(\frac{\widehat{L_i} - L_i}{L_i}\right)^2}{\sum_{i=1}^{n-1}\left(\frac{\widehat{L_{naïve}}_i - L_i}{L_i}\right)^2}} \tag{12}$$

2.5 Building the forecasting model

After collecting all the required data shown in the previous section, we can start building forecasting models. Main motivation for the models are to be simple enough and easy to implement for various countries. We will test the models on the data which we collected for Slovenia and we will compare the results with the existing forecasting model which Slovenian transmission system operator ELES uses. We will use Stata as our main software for all estimation and forecasting procedures. Data sample from 2010 until 2012 will be our estimation period. We will forecast out of sample for next three years and first four months of 2016, from the beginning of 2013 until April 2016.

2.5.1 Single-equation model

First model is the single-equation model. In this model we will estimate all hours in a single equation. We will build it in steps. Reason for that is practical. As already mentioned before, trading companies tend to use derived weather forecasts (for example photovoltaic forecasts, wind production forecasts) instead of direct weather forecasts. By doing modeling in steps, we will be able to see how much each element influences the precision of the forecasts.

Model consist of three different group of variables. First group of variables (X_d), are defying seasonal patterns, second group (X_w), are weather variables and third group (X_l), are lagged demand values. Notation for electricity demand will be L. Simplest formation can be presented in the formula below where each X is a matrix and betas are vectors of coefficients.

$$L = \alpha + \boldsymbol{\beta_d} X_d + \boldsymbol{\beta_w} X_w + \boldsymbol{\beta_l} X_l + \varepsilon \tag{13}$$

As discussed in previous sections, electricity demand has well defined seasonalities. Easiest way to model those seasonalities is by using binary variables for predictors. There should be three to four groups of binary variables:

- first group is for hours in a day,
- second group is for days in a week,
- third group is for months in a year,
- fourth group is for special days.

From the hourly data, we use date and time variable to create binary variables. For the consistency of the data, we use same values as original date and time variables. Hours are defined from 0 to 23, days in a week from 1 to 7, months from 1 to 12. List of holidays in Slovenia can be found on website of The Government of the Republic of Slovenia. Following discussed articles in section 1.3.2.3, we categorize holidays into two groups, depending on the historical effect of the deviation from the normal demand. For example, holidays like Christmas, Easter are considered to be major holidays. Most of the other holidays are considered to be minor holidays. To see how much effect holidays have compared to a normal demand, we create two binary variables, one for major holidays and one for minor holidays. Different specification was used for days in a week interaction. It has three categories, working days, Saturday and Sunday. Reason for that is that not all holidays have the same effect and it also depends on which day in a week is holiday. Additionally, we added day before and day ahead binary variables for holidays to measure an additional effect of bridge days.

Table 3. Coefficients from a regression only on time binary variables

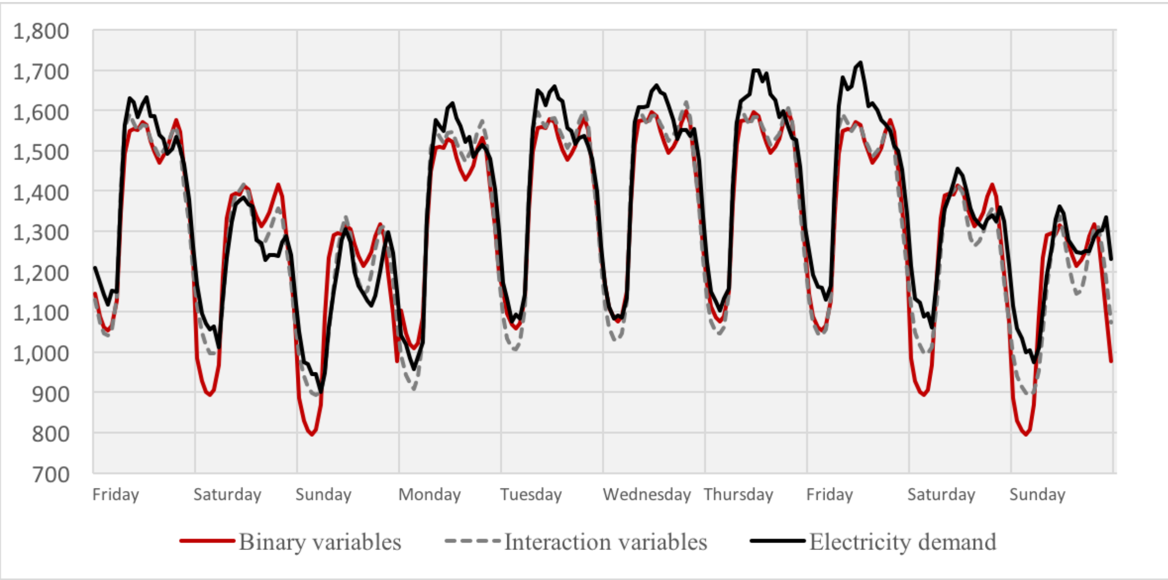
Variable name	**Value**	**Variable name**	**Value**
Constant	1,258.30	Hour 3	-92.59
February	41.77	Hour 4	-80.20
March	-66.54	Hour 5	-17.98
April	-154.75	Hour 6	198.42
May	-160.14	Hour 7	346.56
June	-129.48	Hour 8	404.43
July	-155.71	Hour 9	407.56
August	-187.26	Hour 10	405.41
September	-126.23	Hour 11	426.98
October	-97.25	Hour 12	418.36
November	-64.25	Hour 13	379.95
December	-44.63	Hour 14	351.42
Tuesday	49.14	Hour 15	325.94
Wednesday	65.58	Hour 16	341.30
Thursday	65.74	Hour 17	361.39
Friday	42.85	Hour 18	401.17
Saturday	-117.29	Hour 19	429.56
Sunday	-215.55	Hour 20	400.16
Major holidays	-333.20	Hour 21	308.25
Minor holidays	-322.71	Hour 22	208.47
Hour 1	-56.60	Hour 23	90.48
Hour 2	-83.29		

First we regress demand only on binary variables. Table 3 shows the coefficients from the regression and we can see a difference between days in a week in daily demand compared to Monday. Although Sunday is a non-working day where normal weekly demand is the lowest, demand during holidays is even lower (which we expected). R-squared in this example is 0.82 and all variables are significant on 95% level. To many other areas, this r-squared seems to be quite high, aim in electricity demand modelling is to have an r-squared above 0.90 or even above 0.95.

When plotting the prediction against real values it is easy to notice that there is an interconnection between weekdays and hours. As it can be seen in figure 5 and in figure 14, hourly demand between days in a week is not the same. Solution is to expand binary variables into indicator variables for weekdays and hours (figure 14). In their paper Do et al. (2016) introduce indicator variables for peak hours, for each weekday and each month. That results in 1152 binary variables. In order to reduce a number of variables in a model, one can merge working days or some weekdays as one binary variable. Clements et al. (2016) show that merging weekday variables can produce inferior results. In our case, trying to treat Tuesday, Wednesday and Thursday as one variable (due to similarities between those days) is not better either.

Figure 14 shows two regression results. In first regression electricity demand is regressed only on time binary variables (binary variables, red line). In second regression electricity demand is regressed on interactions variables between hours and weekdays. Although regression on interaction binary variables is an improvement it cannot capture weather sensitive elements. Next step is to add weather variables into the regression.

Figure 14. Line plot sample of binary and interaction variables regressions against real values



Source: *ELES. Load and Generation*, 2016; own calculations.

We will add temperature first. In order to try to explain non-linearity between electricity demand and temperature, a few temperature variables have been created. Fan and Hyndman (2012) suggest using lagged temperatures, for past three hours, and past six days for the same hour. In our model, we use temperature and squared temperature, as well as lagged temperature for past two hours (tests with using more lags did not produce much better results) and for past two days at the same hour. Usage of lagged temperatures Wang, Liu, & Hong (2016) are calling as behavioral "recency effect". It is a psychological aspect where people, for example during a summer period, and after a few hours of similarly high temperatures, one can feel that it is getting warmer each hour. Additional effect of lagged temperatures heating or cooling effect of buildings. In summer periods, if a few days have high temperatures, effect on electricity demand is higher on second or third consecutive day because it takes some time to increase temperature of buildings. Similar scenario can be applied in winter.

Next step is to add an additional weather variable, solar radiation. It also has its seasonal pattern which can be seen in figure 15, although it is not that strong when we compare it to the seasonality effect of the temperature.

Figure 15. Scatter plot between electricity demand and solar radiation (for working days, by months)

Source: *ELES. Load and Generation*, 2016; *Ministry of the environment and spatial agency - Slovenian Environment Agency*, 2016; own calculations.

Table 4 shows the error measurements of the regressions. For this example, lagged realized demand is not included. Normal scenario is model only with indicator binary variables, temperature scenario is normal with added temperature variables, and solar scenario is final scenario where we added solar variable in temperature scenario.

Table 4. Error measurements of predictions for each new variable added in regression

	Error values
MAE Normal	61
MAE Temperature	52
MAE Solar	51
MSE Normal	1,320
MSE Temperature	1,329
MSE Solar	1,329
MAPE Normal (in %)	4.53
MAPE Temperature (in %)	3.91
MAPE Solar (in %)	3.87

For normal scenario errors are largest and MAPE is 4.53%. We already know that temperature has a non-linear effect on demand, and by adding temperature functions demand errors significantly decrease for more than a half percent. Errors only slightly changes downward if we add a solar variable into a regression. That is why temperature is most used weather factor in modeling forecasting electricity demand models.

After adding indicator variables for modelling seasonalities and weather variables, next step is to add lagged realized demand. Since we are interested only in direct forecasting, for day-ahead forecasting, first available lag for the same hour is 48 hours before. Also, we add one more lagged realization, 7 days before for the same hour (168th hour lag). There are two reasons for preference of direct forecasting compared to multi-step forecasting. First reason is that in theory, direct forecasting models are more robust to the model misspecification comparing to the multistep iterative forecasts (Marcellino, Stock, & Watson, 2006; Taieb & Hyndman, 2012). Second reason is reason is the speed of the whole process. If a trading company currently is interested only in day-ahead demand (which can be the case), then direct forecasting model is faster to implement.

In their paper Clements et al. (2016) discuss possible issues with using lagged working days to predict Saturdays and Sundays or lagged Saturdays and Sundays to predict Mondays. Because during workdays demand is generally greater than during the weekends, it leads to over-prediction of weekends (negative bias in the errors) and under-prediction of Mondays (positive bias in the errors). In their case reason for this is that the coefficients on one-day (they are using 24th hour lagged demand) lagged demand do not differentiate between days of the week. Solution for this issue is to let lagged values interact with day of the week binary variables.

Final formula is for modelling electricity demand is

$$Demand = \alpha + \boldsymbol{\beta_1} IND_{time\,(h,d,m)} + \boldsymbol{\beta_2} IND_{hol} + \boldsymbol{\beta_3} TEMP + \beta_4 Solar + \boldsymbol{\beta_5} Lag\ demand \cdot IND_{time(d)} \tag{14}$$

2.5.2 Multi-equation model

Different approach to modeling hourly electricity demand is to treat each hour as individual time series. Base for this model is the paper from Ramanathan et al. (1997), with which they won the electricity demand forecasting competition. For the comparison purposes, we define variables in the model as close as possible to the single-equation model.

Main difference between our model and the one proposed by Ramanathan et al. (1997) is, that in default scenario, all weekdays are included in the model, so the end model has 24 equations, instead of 48. Also, we will not use lagged errors in the model.

One practical reason which could be in favor of building a model for individual hours is that one does not need to account for complicated interactions between hours and days in a week. Direct consequence of that is less variables in each regression and the ability for older (slower) computers to process the related regressions. Also, there is a possibility that multi-equation models can reduce noise in the data.

Formula for modeling hourly demand is:

$$Demand_h = \alpha + \boldsymbol{\beta_1} IND_{time\,(d,m)} + \boldsymbol{\beta_2} IND_{hol} + \boldsymbol{\beta_3} TEMP + \beta_4 Solar + \boldsymbol{\beta_5} Lag\ demand \cdot IND_{time(d)} \tag{15}$$

Both models are estimated using OLS. Although by using lagged dependent variables in a OLS regression, assumption of strict exogeneity does not hold. Instead, assumption of weak dependence can be used, and in order that this assumption is satisfied, stability condition ($|\beta| < 1$) must also be satisfied. Estimated coefficients are biased, when sample size is small. In case one has a large sample, estimated beta should be a good estimator of true beta. Apart from alternatives, OLS regression in forecasting electricity demand is quite popular (Ramanathan et al., 1997) and we have a large sample, so we decided to use OLS regression as well.

3 EMPIRICAL RESULTS

In the previous chapter we explained data structure and we built forecasting models. We kept models' specifications are as close as possible to each other, that we can easily compare them. In this section we will first compare models to try to see if there are any differences in forecasting results and also, what would be the possible reason for those differences. After that we will introduce benchmark and adjust the models so the final model can be comparable to the benchmark.

In the end, we will investigate whether combining models would produce better forecasting results compared to the individual models.

3.1 Comparison of the models

For the comparison purposes, both models have the same specification. Ramanathan et al. (1997) give an additional advantage to multi-equation models by having the ability of making different specifications for each hour, since they expect, that different hours have different sensitivities to external factors and using the ability to model each hour independently could give an additional advantage to multi-equation models. For our comparison, we will keep the specifications of the models the same. Models will be compared in-sample and out-of-sample.

In table 5 and table 6 we present the results of the models for in-sample and out-of-sample periods (hourly frequency). Accuracy of any forecasting model depends on the data on which the model is specified. It is expected that out-of-sample results are not as good in comparison with the estimation period. In our case, average change is not too big. MAPE increased for almost half a percent. Out-of-sample change would have been smaller if we had modeled the single-equation model without interactions between months and weekdays, which would in turn made our overall error larger (for approximately 0.3-0.4%).

Tables are indicating that multi-equation model is slightly better, but we would like to also analyze forecasting errors (primarily out-of-sample) structured by hours in a day, days in a week and by months. That could give us a better understanding of the reason why the multi equation model is slightly better (since both models have the same specifications).

Table 5. Error measurements between models (with 48-hour lag) comparison with included holidays

	In-sample errors	**Out-of-sample errors**
MAE Single	37	44
MAE Multi	36	43
MSE Single	2,533	3,457
MSE Multi	2,449	3,347
MAPE Single (in %)	2.80	3.27
MAPE Multi (in %)	2.73	3.19

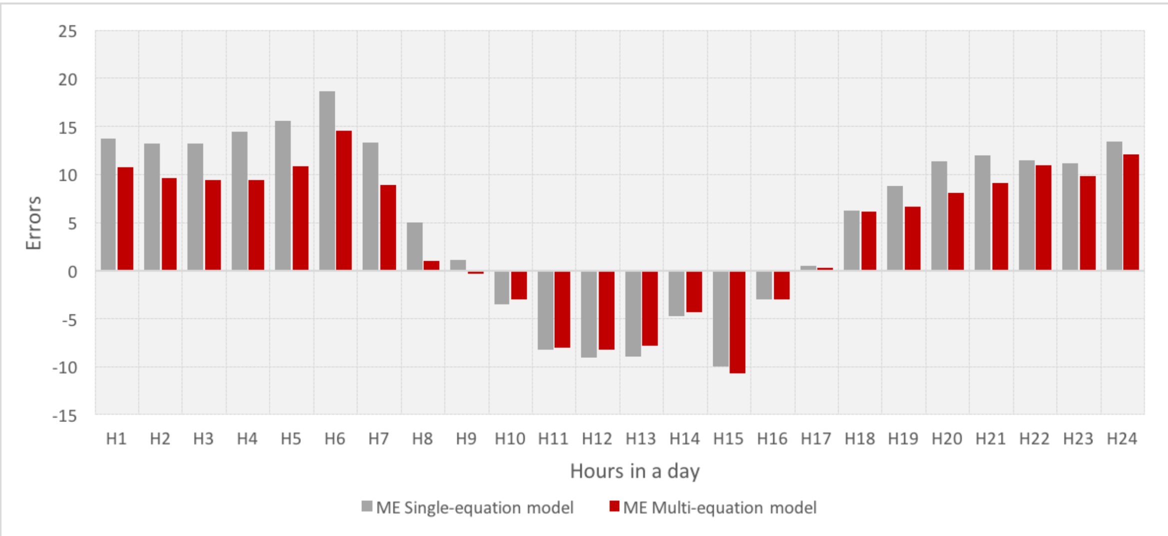
Table 6. Error measurements between models (with 48-hour lag) comparison without holidays

	In-sample errors	Out-of-sample errors
MAE Single	35	42
MAE Multi	34	41
MSE Single	2,224	3,029
MSE Multi	2,134	2,840
MAPE Single (in %)	2.61	3.05
MAPE Multi (in %)	2.53	2.95

3.1.1 Mean error

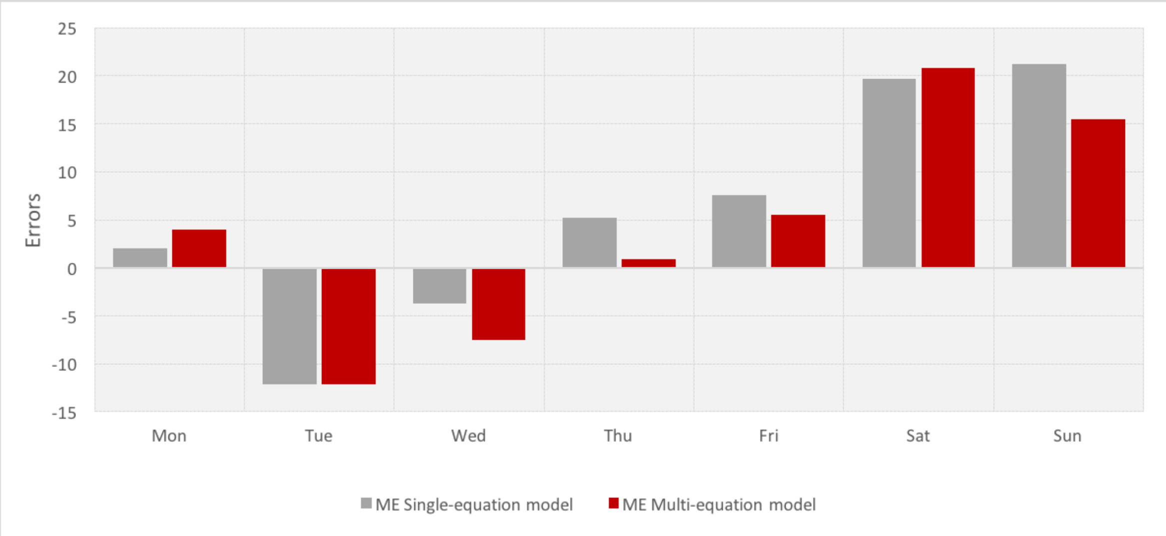
Figures 16, 17 and 18 show mean errors of both models. Vertical axes show mean errors in MWh and horizontal axes show hours of the day (from 1 to 24), days in a week (from 1 to 7) and months in a year (from 1 to 12) respectively. In figure 16 we can see out-of-sample period mean errors of single- and multi-equation models. Mean errors are showing us direction of the errors. In most hours, both models are underestimating, but in peak hours, where demand is highest, models are overestimating. In both cases, multi-equation model has lower error than single-equation model. Only hour 15 has a larger error. Because mean error is simple average of positive and negative values, it can only tell us direction of the errors. What is also noticeable that the off-peak hours are much less underestimated in the multi-equation model.

Figure 16. Out-of-sample ME of single- and multi-equation models for hours in a day



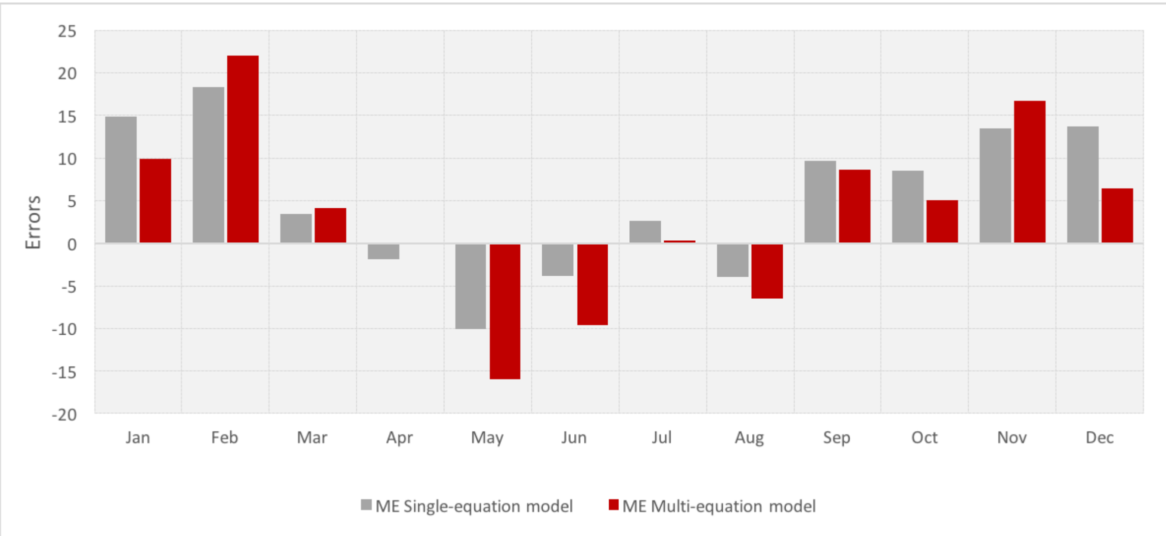
On a daily basis, mean errors in both models show that working days are better estimated than weekends, due to the fact that errors in working days mostly cancel out compared to weekends. Interesting is, that Tuesday is quite overestimated in comparison to other working days. Also, weekends are on average underestimated, where multi-equation model is more suitable from the mean error perspective.

Figure 17. Out-of-sample ME of single- and multi-equation models for weekdays



Monthly perspective shows similar pattern as the hourly one. In winter periods forecasts are mostly underestimating, while in other months under or overestimation is lesser, while in spring and most of the summer periods models are overestimating. Months with smallest average errors for both of the models are April and July.

Figure 18. Out-of-sample ME of single- and multi-equation models for months

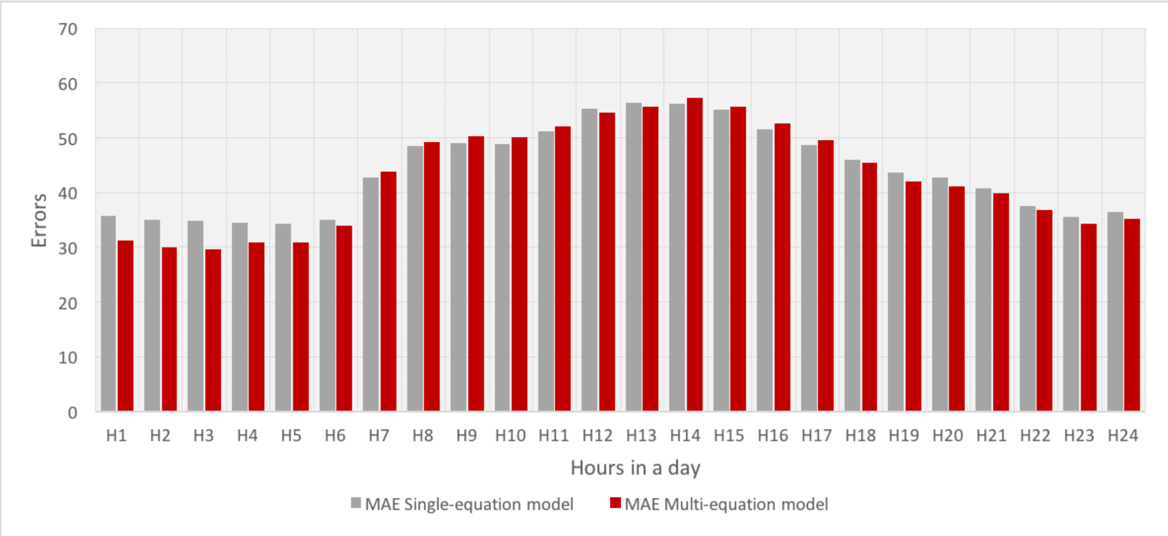


3.1.2 Mean absolute error

Figures 19, 20 and 21 show mean absolute errors of both models. Vertical axes show mean absolute errors in MWh and horizontal axes show hours of the day (from 1 to 24), days in a week (from 1 to 7) and months in a year (from 1 to 12) respectively. Absolute errors are showing how much in nominal values forecast deviates from the real values. Since peak hours are higher in comparing to off-peak hours, it is logical to assume that errors will be larger in absolute terms. Reason for that is that sensitivity (and with that volatility) of those hours is higher during the peak hours.

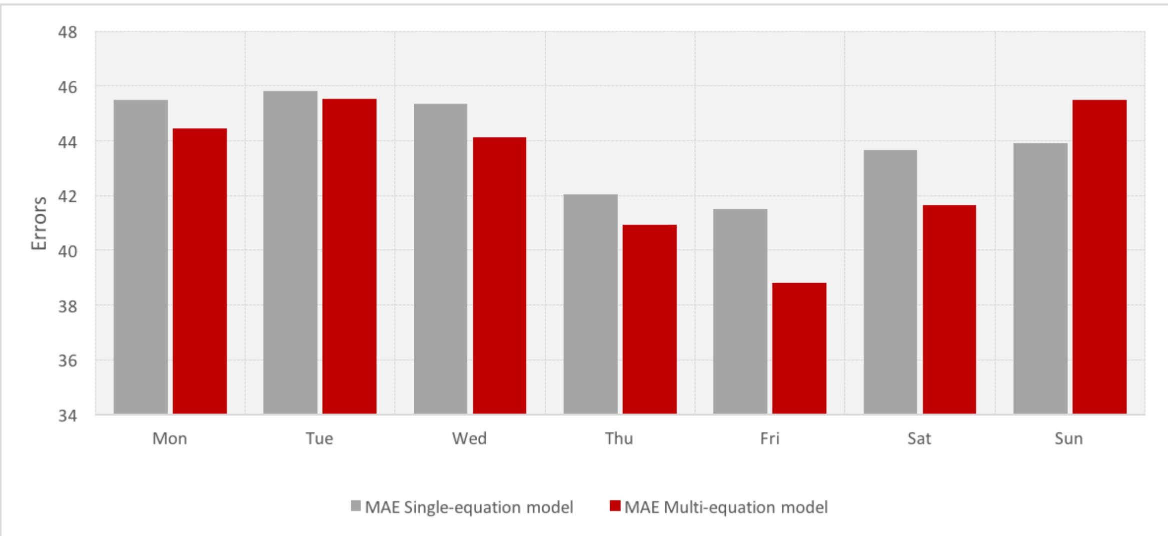
Both models are quite similar regarding the MAE perspective. Main difference is that night hours have larger absolute error for the single-equation model. This could be the main reason for better overall results of the model.

Figure 19. Out-of-sample MAE of single- and multi-equation models for hours in a day



According to a daily perspective, for both models smallest errors are on Thursdays and Fridays. Multi-equation model is only slightly worse on Sundays. Largest difference is on Friday and also multi-equation model has the smallest error. Errors between working days are different, although days from Monday to Wednesday are on a similar level. This could also be the reason why attempting to reduce the number of variables by merging some of the weekdays results in a larger error.

Figure 20. Out-of-sample MAE of single- and multi-equation models for days in a week



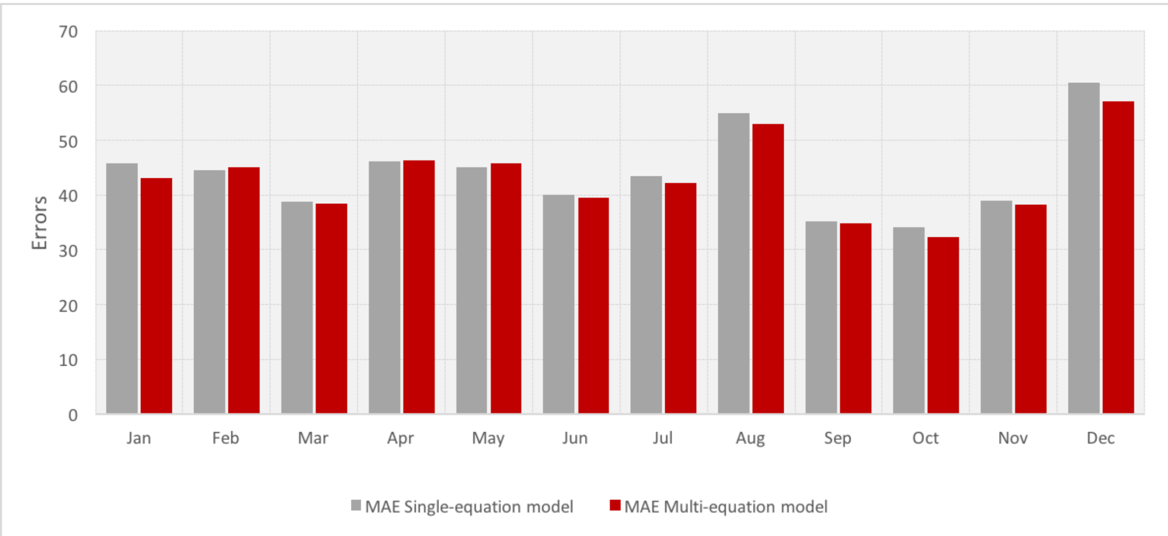
Monthly absolute errors are quite similar for the first seven months. We can see few spikes in two months, August and December. Although we have monthly binary variables (or interactions) in our models, most likely there two scenarios which could explains this change.

There is an additional unobserved holiday effect in a certain part of the month. In August, for example, lots of people are going abroad for vacations. Since mean errors are low for August, that means that some of the errors are canceling each other. Monthly binary/indicator variable captures average effect (if we ignore weather factors, we defined monthly effect as a binary variable) for the whole month and because of that it cannot capture an unobserved effect or that effect influences the level for the whole month.

Similar explanation can be applied to December. During winter holidays (Christmas and New Year) people are usually taking vacations, so instead of few days of official holiday, demand could be much lower for a whole week or two. In this case, models are underestimating the December month, probably because holiday effect here is much stronger so it lowers the whole month.

Other or additional scenario for December could be interpreted if we look at the neighboring months in figure 18. Forecasts for neighboring months are also underestimating. In addition to holiday factor, it could mean that some possible unobserved weather factors could have some influence. One example could be wind which in cold temperatures “helps” buildings cool faster which increases demand for heating and, in the end demand for the electricity. Also there could be some behavioral or industrial factor (for example, irrational use of heating in winter months).

Figure 21. Out-of-sample MAE of single- and multi-equation models for months

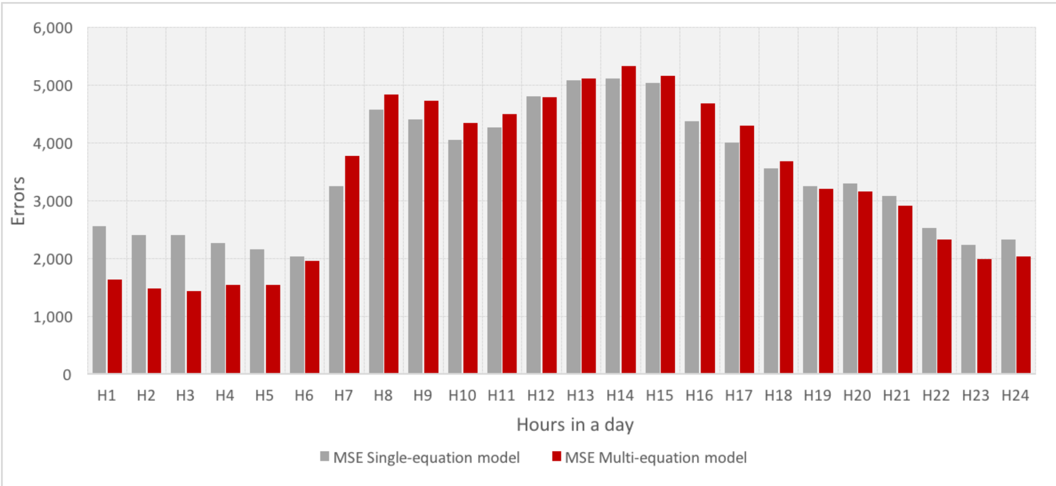


3.1.3 Mean squared error

Mean squared error is just an additional confirmation of mean absolute error. Figures 22, 23 and 24 show mean squared errors of both models. Vertical axes show mean squared errors in MWh and horizontal axes show hours of the day (from 1 to 24), days in a week (from 1 to 7) and months in a year (from 1 to 12) respectively. Errors seen during the night hours are larger in single-equation model than in the multi-equation one. What is interesting to see and that is not so noticeable in figure 19 is that mid peak hours are slightly better estimated by the single-equation model. There are additional two things which can be seen in figure 22.

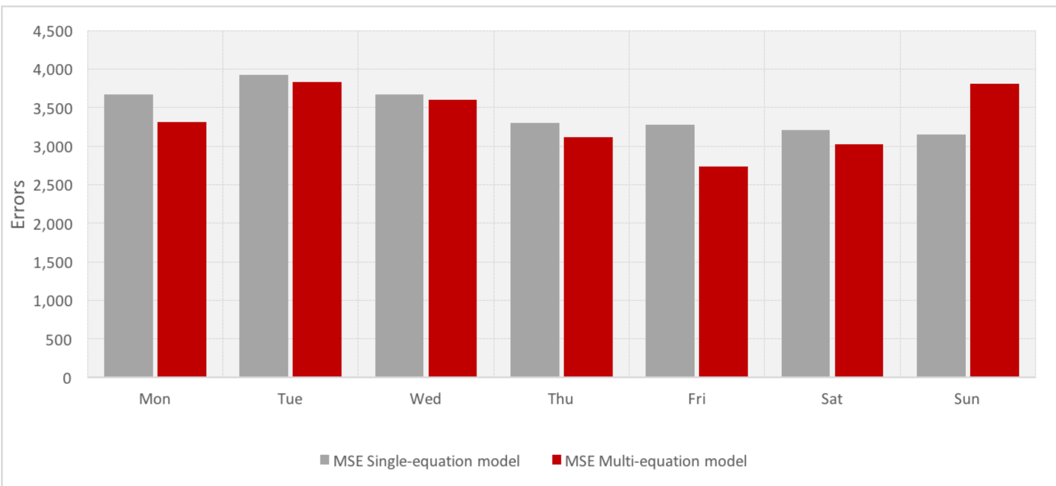
At around seven in the morning, as the activity of the population increase, so do the errors of both models (we explained the hourly normal "shape" in section 1.2.1.1). Second errors jump happens in the afternoon, when activity temporarily drops (when people are traveling home from work, or resting just after work). At the second increase of demand errors are lower.

Figure 22. Out-of-sample MSE of single- and multi-equation models for hours in a day



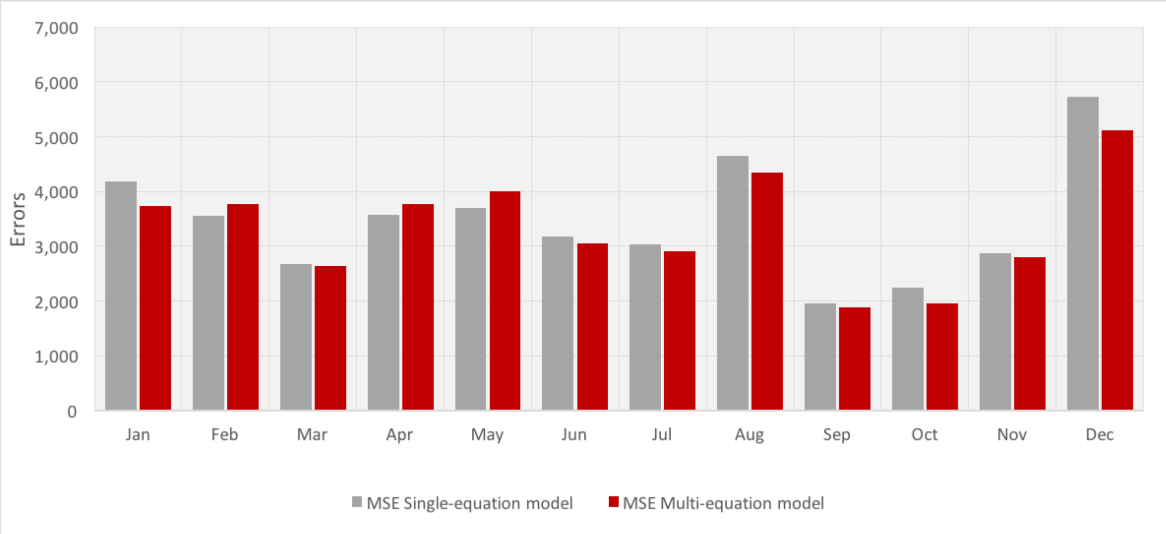
Possible explanation for better peak estimation of single-equation model lies in figure 23. If we check the days in a week again, largest difference between forecast is on Sunday where multi-equation model is worse. This was also seen on MAE charts, but since MSE "penalizes" larger errors by squaring them, here it is clearer to see the possible reason for larger peak errors.

Figure 23. Out-of-sample MSE of single- and multi-equation models for days in a week



Mean squared errors on a monthly basis is not showing different error patterns from mean absolute errors. Lower values in September and October are confirming that those two months are the most stable ones.

Figure 24. Out-of-sample MSE of single- and multi-equation models for months

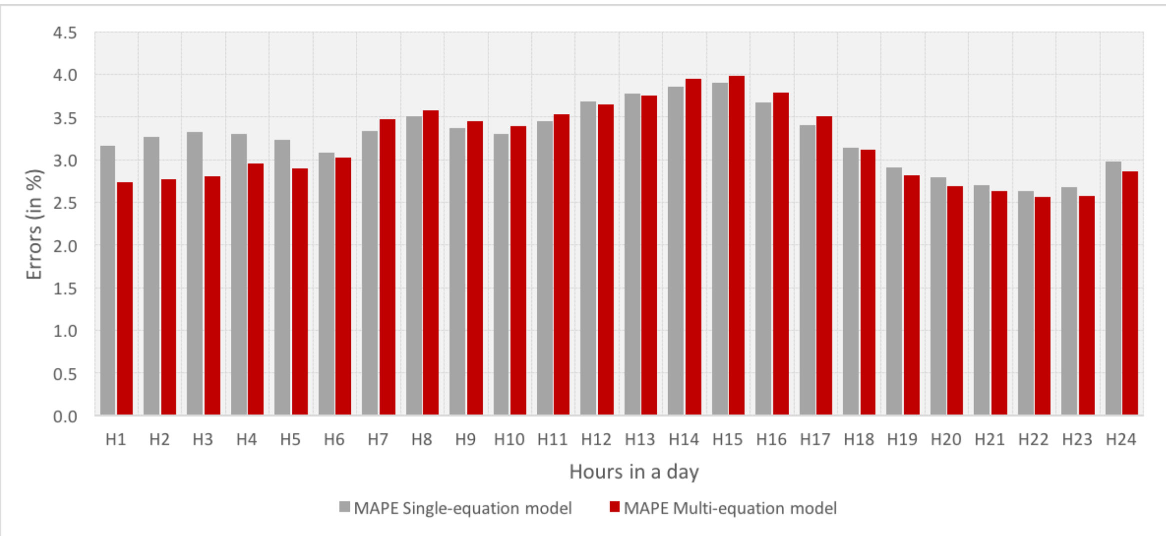


3.1.4 Mean absolute percentage error

Figures 25, 26 and 27 show mean absolute percentage errors of both models. Vertical axes show errors in percentage points and horizontal axes show hours of the day (from 1 to 24), days in a week (from 1 to 7) and months in a year (from 1 to 12) respectively.

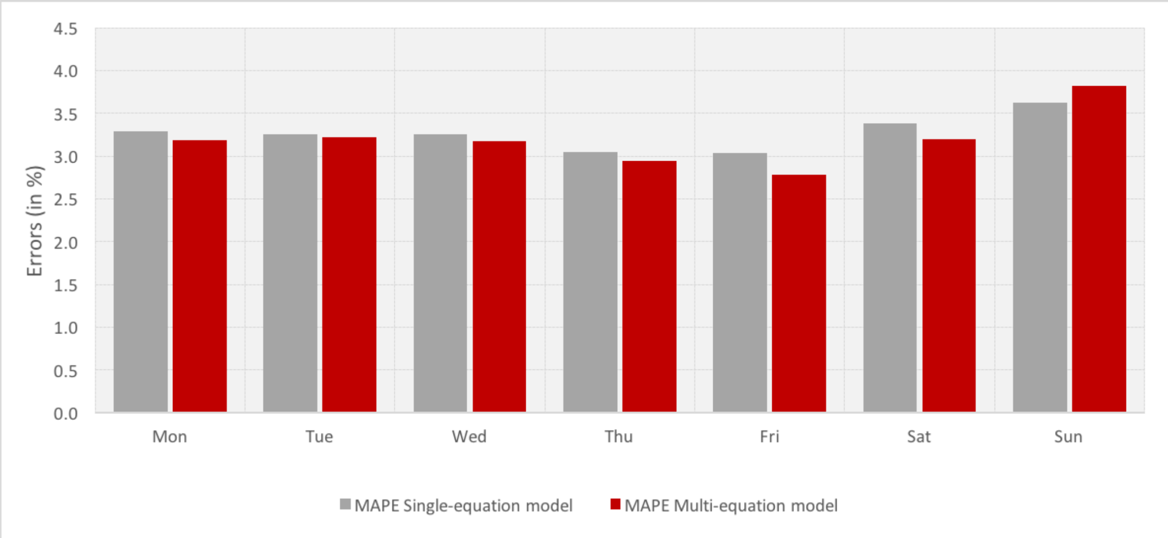
Figure 25 shows the hourly mean absolute percentage errors. During the night hours when population activity is at its lowest level, MAPE is also lower for the multi-equation model. Last hour error is also larger for multi-equation model.

Figure 25. Out-of-sample MAPE between single- and multi-equation model for hours in a day



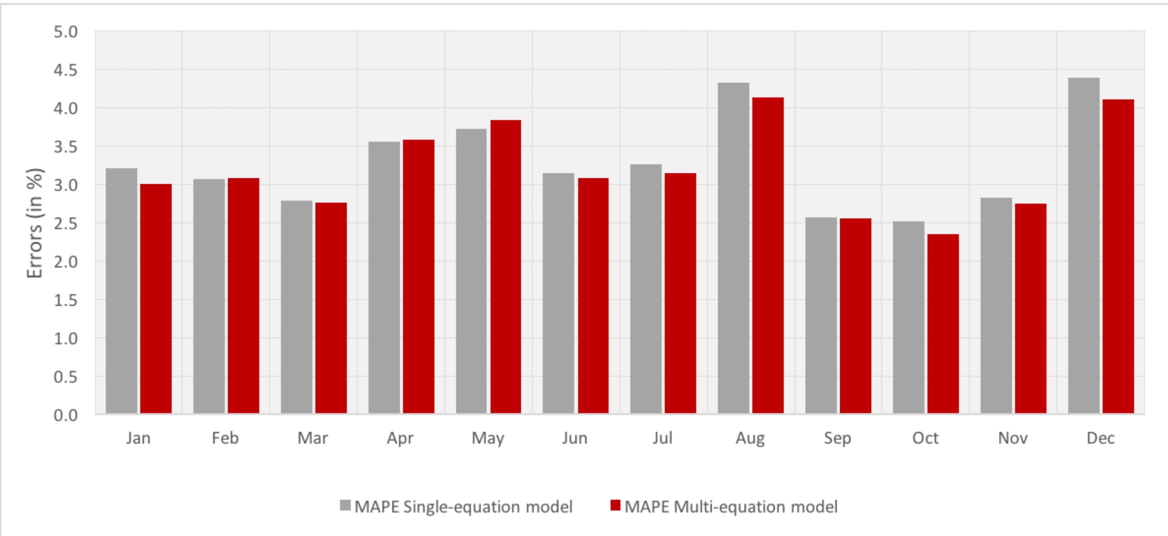
MAPE also shows one interesting thing. Weekends have larger percentage errors than working days. That cannot be seen on absolute error charts since errors are shown in absolute values, and demand on weekends is lower than demand on working days. Also, from the MAPE perspective, Sunday has the largest error, absolute and in percentage points.

Figure 26. Out-of-sample MAPE between single- and multi-equation model for days in a week



On the relative scale of errors, large errors for August and December are also confirmed. Additionally, errors in April and May are relatively larger in comparison to other months in first half-year. In absolute terms that was not the case.

Figure 27. Out-of-sample MAPE between single- and multi-equation model for months in a year



3.2 Choosing the final model

After checking all errors measurements as well as trying to understand and explain the behavior of the model on different time factors (hour of the day, days in a week, months in a year), multi-equation model is picked to be better model. There are three main reasons for this:

- with the same specifications of the models, multi-equation model has slightly better results,
- model is more flexible because it gives an option to model each hour differently in order to get even better results,
- since there are less variables in a multi-equation model per equation (and less data for each hour), it is easier to estimate and interpret the results.

It is also worth noting that this is not the universal case. It could be possible that for some other country, a single-equation model would be better than a multi-equation one. Development of technology gives the analysts the possibility to do more tests faster and in situations when one is using only one technique for building a forecasting model, which is why we advise testing both variations.

3.3 Replicating the benchmark

Now that we have picked the final model, we would like to know how good it is when compared to the benchmark. There are lots of articles where benchmark is usually some simpler model, like the naïve approach. In the paper, Wang et al. (2016) discuss that using naïve approach for benchmark is not a good idea, since naïve models are usually quite imprecise. Naïve approach in this case would be considering Monday from previous week or some weighted average of a few previous Mondays as forecast for this Monday or some similar approach.

In the case of Slovenia, we do not have to build a benchmark model since there are already available historical forecasts from Slovenian TSO. We will use their forecasts as a benchmark. As mentioned before in section 2.1, ELES has different leading times for publishing forecasts comparing to our models. To be able to compare our model with a benchmark, we have to make some additional variations of our model.

As shortly mentioned before, ELES publishes their forecasts for day-ahead with 14 hours of leading time from Tuesday to Saturday. For Sunday and Monday leading time approximately increases to 32 and 56 hours respectively (they publish forecasts for Sunday and Monday after 4 p.m. on Friday). Replicating forecasting model is a combination of two direct forecasting models (variation of a single model by picking first available lagged variable).

Our model is built with 24 hours of leading time. That means, at midnight today, model is forecasting hourly demand for tomorrow. When ELES publishes their forecast, approximately after ten in the morning, they already have the possibility of using information from realizations for first ten hours of today. That gives them a certain advantage. In order to replicate the benchmark, we will create an additional model which in its structure has the same minimal lag structure.

Since most of the trading companies are working every day increasing minimal lag structure (or doing multistep forecasting) for Sunday and Monday is not necessary and we will ignore those days in the analysis.

Replicated and day ahead models will be compared and analyzed in a few ways. We are interested how they both handle seasonalities, so we will compare models the same way as we tested our initial models (by hours, by days in a week and by months). Also, we want to see how much special days have influence on the result.

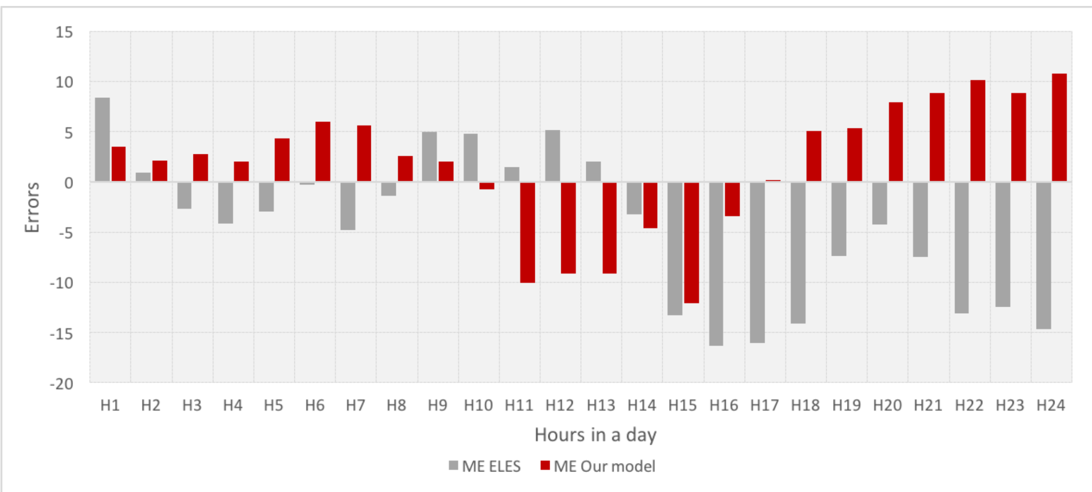
After comparing models, we would like to see if there is any additional effect on combining benchmark with our forecast in order to try to improve forecast accuracy. Belief that combining forecasts will improve their accuracy is not a new approach. It was investigated before and there is a number of articles on this topic. One of those articles are from Clemen (1989) and Nowotarski, Liu, Weron, & Hong (2016).

We will combine forecasts in two ways: one way would be to regress the forecasts using OLS; other way would be to use the average.

3.3.1 Mean error of the replicating procedure

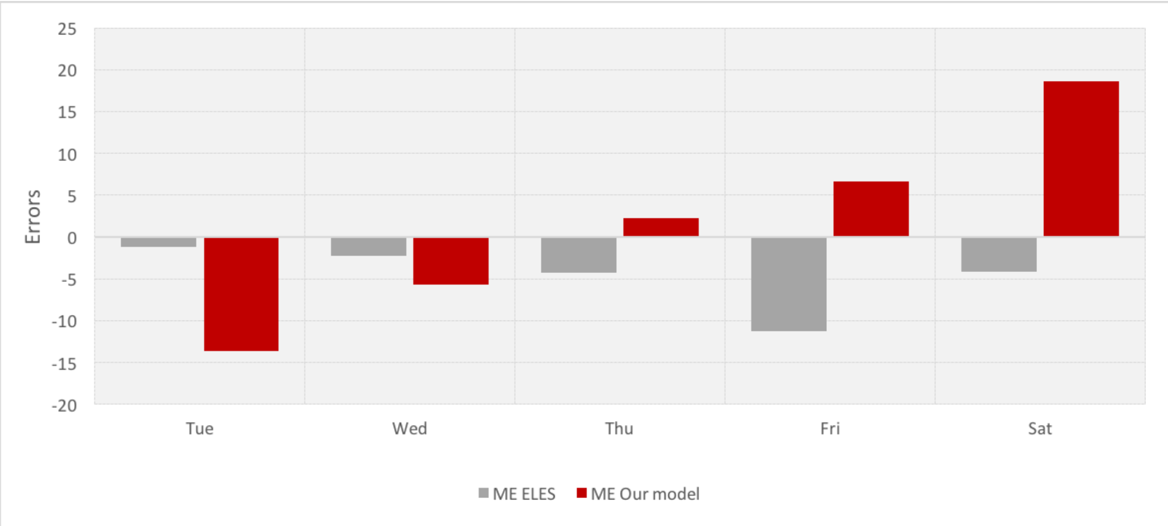
Benchmark's model has a different hour mean error structure (figure 28). For the first part of the day, errors are much lower compared to the afternoon where their model on average is quite overestimating. In the hours of second "spike" during the day in the demand, the benchmark model better captures the lack that jump. Our model is better using information from change in minimal lag structure and for first ten hours decreased errors.

Figure 28. Out-of-sample ME comparison between benchmark and our model for hour in a day (Tuesday-Saturday)



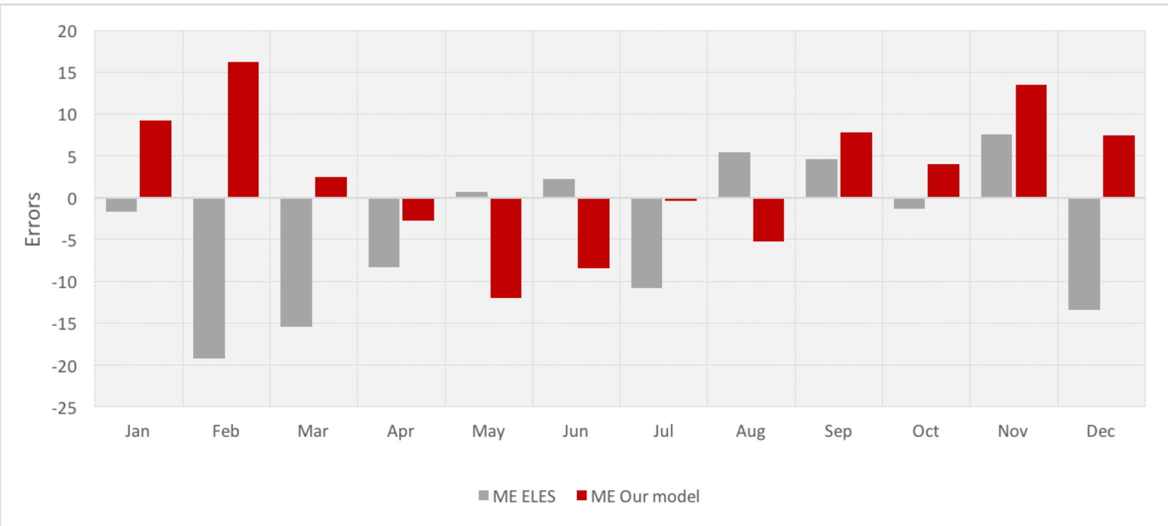
Using same day-ahead forecasting structure, we can only compare days from Tuesday to Saturday. In days in a week case (figure 29), benchmark model behaves better, more stable. In benchmark model Fridays are much more overestimated than other days. Our model is overestimating Tuesdays and underestimating Saturdays. Comparing both models mean errors on average (on daily basis) are in different directions.

Figure 29. Out-of-sample ME comparison between benchmark and our model for days in a week (Tuesday-Saturday)



On a monthly basis both of the models have similar mean error but in most of the cases they also have opposite signs. For the benchmark model much larger errors are seen in February, March, July and December, and much smaller errors are in January, June and October.

Figure 30. Out-of-sample ME comparison between benchmark and our model for months in a year (Tuesday-Saturday)

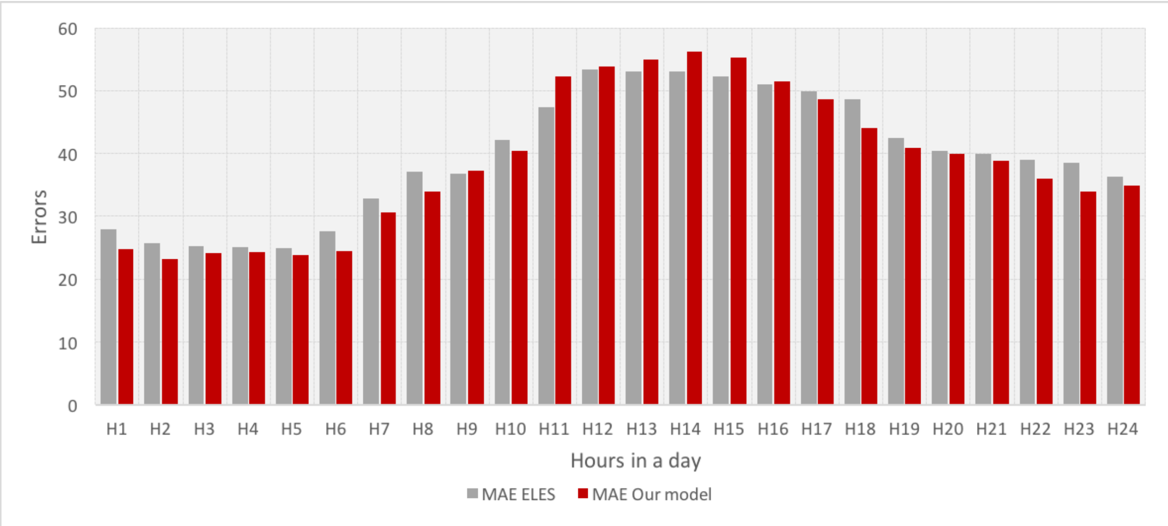


3.3.2 Mean absolute error of the replicating procedure

Shapes of the errors for both models are almost the same. There are two interesting patterns which can be noticed in figure 31:

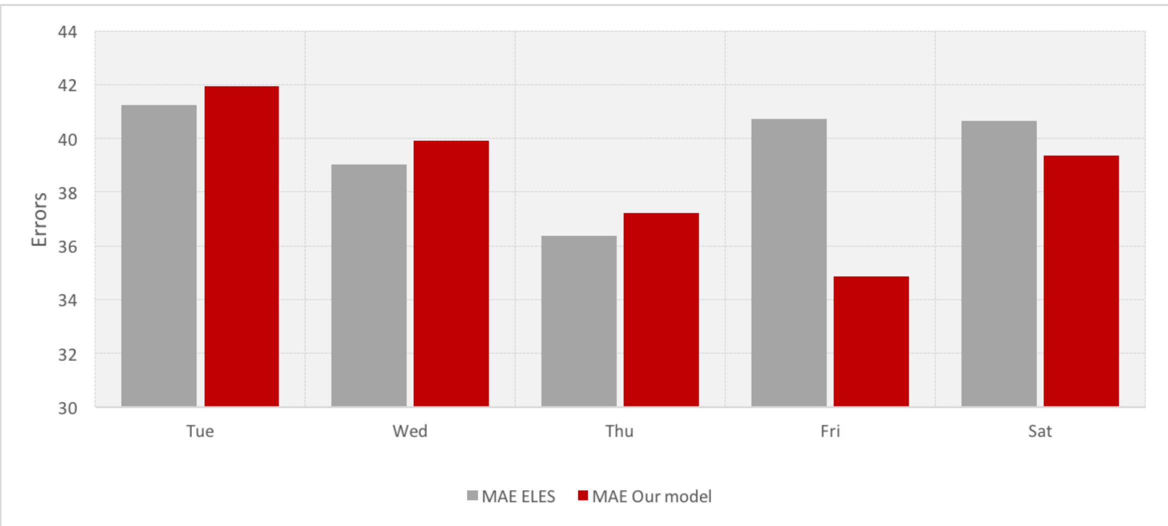
- proposed model is slightly better using information for night hours,
- in mid peak hours, the benchmark model has slightly smaller errors, but after hour 17 proposed model picks up (apart from hour 20).

Figure 31. Out-of-sample MAE comparison between benchmark and our model for hours in a day (Tuesday-Saturday)



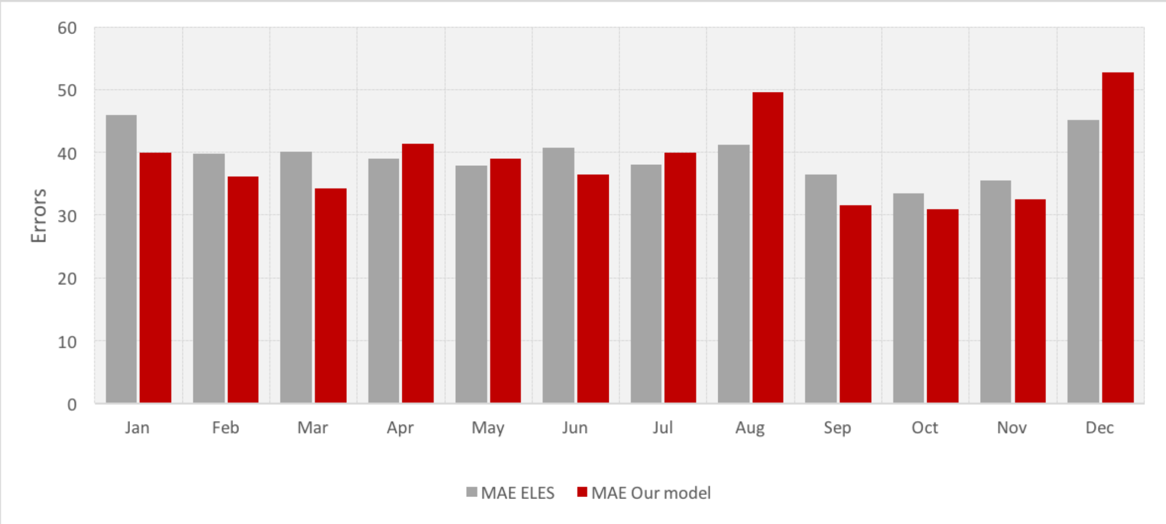
Errors of our model are greater for Tuesday and Wednesday, slightly worse for Thursday, but errors for Friday are quite large for benchmark model comparing to ours. Saturdays' errors are similar (slightly in favor of our model).

Figure 32. Out-of-sample MAE comparison between benchmark and our model for days in a week (Tuesday-Saturday)



Mean absolute errors are showing less monthly volatility in the benchmark model, where our model has spikes in August and December. Possible reasons for those errors are discussed in section 3.1.2. For other months our model shows better results, only in April and May, benchmark model has less mean absolute error.

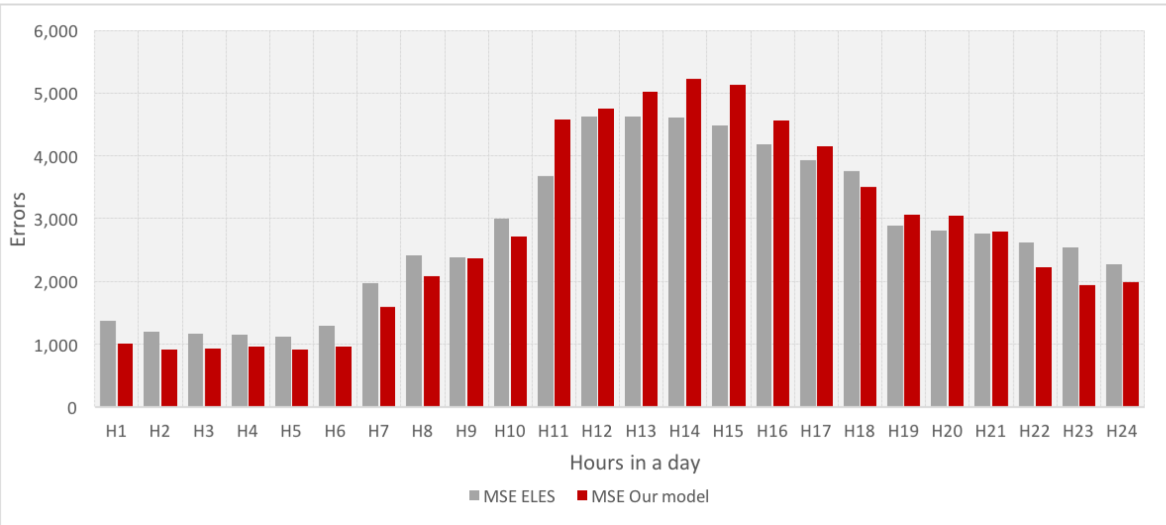
Figure 33. Out-of-sample MAE comparison between benchmark and our model for months in a year (Tuesday-Saturday)



3.3.3 Mean squared error of the replicating procedure

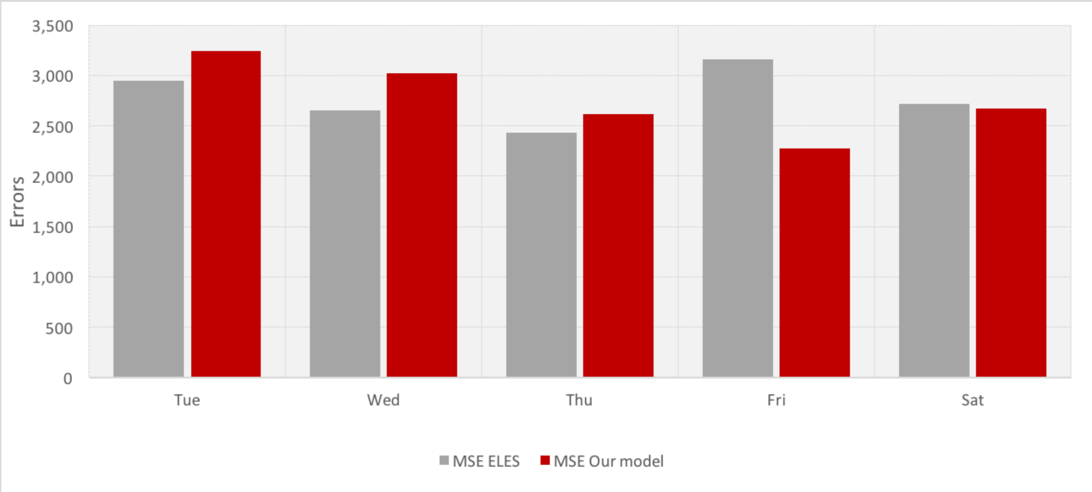
Since MSE penalizes large errors, from the hourly perspective, it confirms mid peak hours mean absolute error results. Since electricity demand in night hours are by default lower, changes in mean squared errors should not be much that higher. Still, the differences between forecasts in night hours are also greater.

Figure 34. Out-of-sample MSE comparison between benchmark and our model for hours in a day (Tuesday-Saturday)



Both models behave similarly, but our model is much more stable on Fridays. For Tuesdays and Wednesdays, benchmark model shows better results.

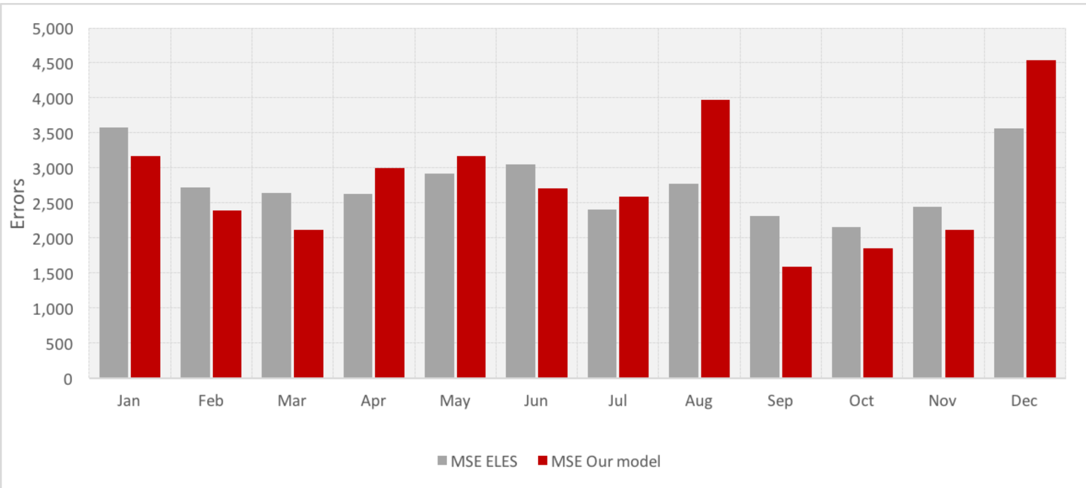
Figure 35. Out-of-sample MSE comparison between benchmark and our model for days in a week (Tuesday-Saturday)



On a monthly basis, benchmark model does not have spikes during summer. August months have similar errors to other months. This confirms that August has some unobserved effect on demand. Possible approach to solving this issue is to try to model monthly effects. If there is some holiday effect, in the beginning of August, demand should be lower disregarding other factors. One could create an additional kind of category for holidays which could help to identify unobserved August effect and adjust the demand for those specific periods if assumption of the error is correct.

For both models holiday effect in December makes larger average errors. Difference between months in our case is much larger compared to the benchmark model, but also errors of our model are lower for previous month and also for most of the other months.

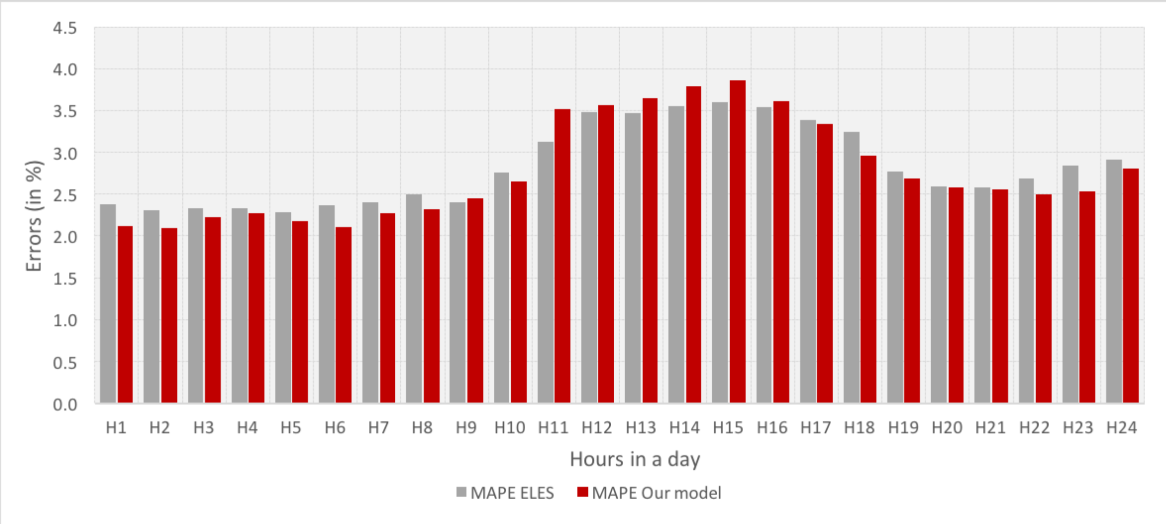
Figure 36. Out-of-sample MSE comparison between benchmark and our model for months in a year (Tuesday-Saturday)



3.3.4 Mean absolute percentage error of the replicating procedure

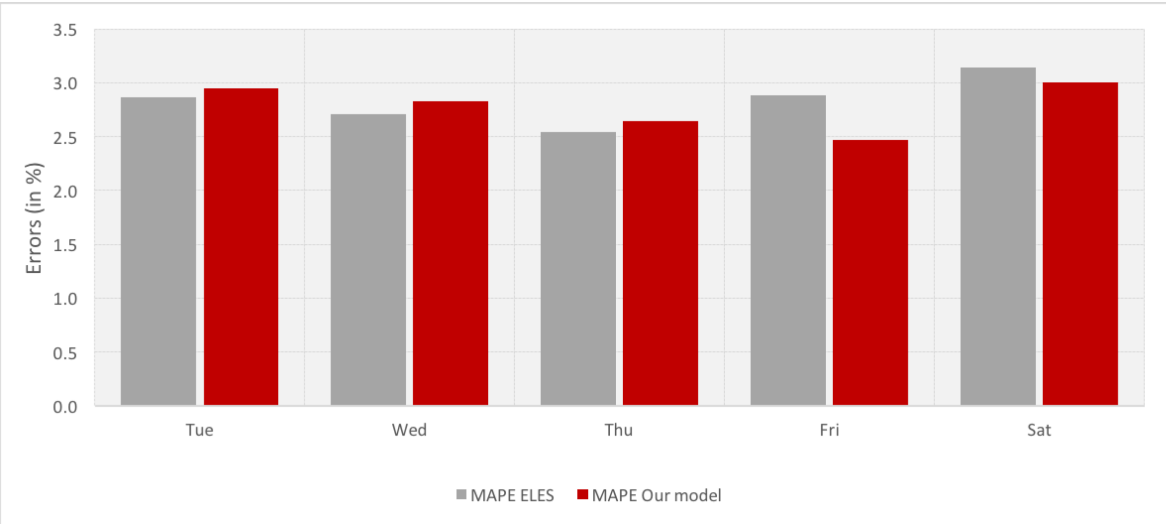
MAPE on hourly basis does not say anything more than MAE and MSE. Our model is better in off-peak hours while in mid peak hours the benchmark model has better results.

Figure 37. Out-of-sample MAPE comparison between benchmark and our model for hours in a day (Tuesday-Saturday)



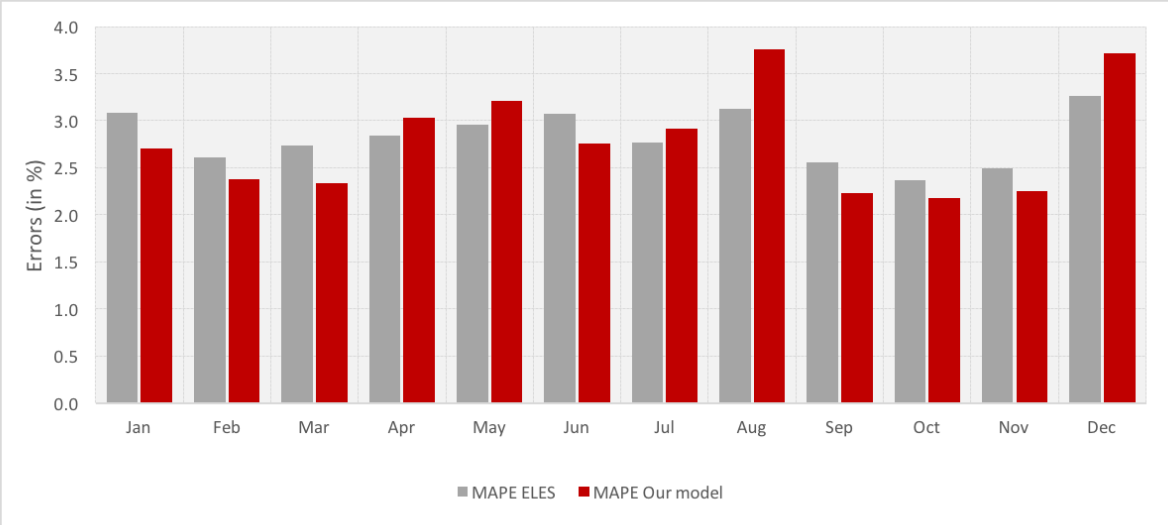
On daily basis results are also consistent with two previous measurement errors. Our model better captures Friday effect (on Friday demand in the afternoon is lower when compared to other working days).

Figure 38. Out-of-sample MAPE comparison between benchmark and our model for days in a week (Tuesday-Saturday)



MAPE on monthly basis is also similar like MAE and MSE, but there is an additional information on this chart. We already suspect that our model has some potential misspecification regarding unobserved effects in August and December, but what is also interesting is that highest errors in relative terms for the benchmark model are also in those two months. That only confirms that this effect cannot be completely captured with standard time factors and one should think about adding some additional variables.

Figure 39. Out-of-sample MAPE comparison between benchmark and our model for months in a year (Tuesday-Saturday)



3.4 A short summary of the analysis of measurement errors

After presenting errors for all seasonalities and comparing them with the benchmark, some conclusions are to be made. Models are compared only for day-ahead forecasts, from Tuesday to Saturday.

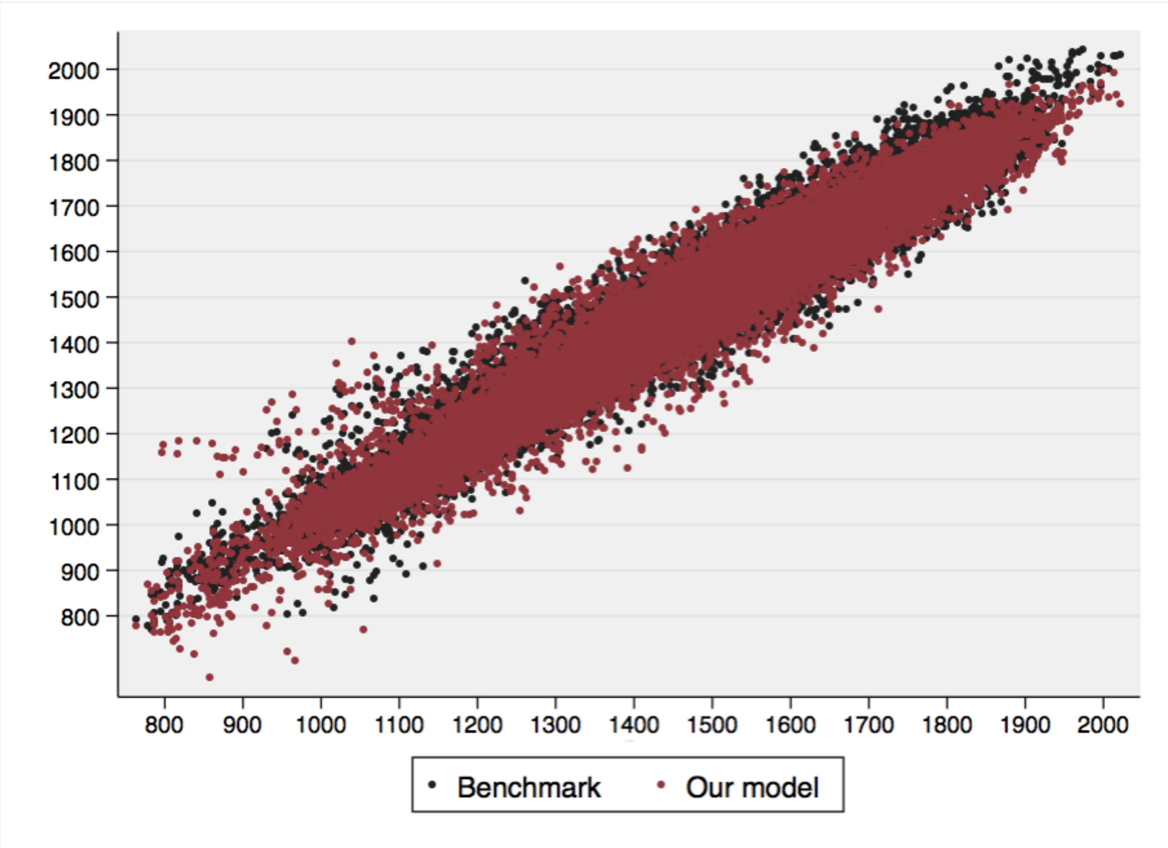
Firstly, on an hourly basis, our model is much better for the first hours. It better uses information gained from changed minimum lag structure and in those hours all error measurements are showing better results. In mid peak hours, the benchmark model has a slight advantage and also variance of the errors in our model is larger in second jump in the demand. Other hours are quite similar to our model or our model has some advantage. On a daily basis models are similar, but there is a larger difference on Friday. Difference is larger in absolute terms than in relative terms. Also average errors are mostly in opposite direction. That would mean that on average, errors could cancel each other if models are combined. Absolute errors on monthly basis are confirming that there is some unobservable effect in some months which cannot be captured by standard variables.

Finally, we can look at the scatterplot between the benchmark and our model in figure 40. Dark red dots represent proposed model, and black dots represent the benchmark model. On the edges of our surface we are able to see black dots from the benchmark model. There is not much difference between the shapes or the density of both models. It is easy to see that models are quite similar. Visible difference lies in some outliers in the lower region of the chart. That suggests that benchmark model better handles special days or bridge days because lower values of the demand suggest at special days or weekends, but since special days are rare events it is more likely that forecast errors are larger than for the weekends. Table 7 confirms that we have successfully replicated a benchmark model, though, we were unable to significantly improve it. Also, errors are smaller in a sample which does not include holidays. Theil U-statistic is almost equal to 1 when holidays are included, and it is around 0.95 without holidays.

Table 7. Results of the replicating procedure (Tuesday-Saturday)

	With holidays	Without holidays
MAE Benchmark	40	39
MAE Our model	39	37
MSE Benchmark	2,783	2,630
MSE Our model	2,766	2,413
MAPE Benchmark (in %)	2.826	2.712
MAPE Our model (in %)	2.777	2.592
Theil U-statistic	1.023	0.951

Figure 40. Scatterplot between benchmark (dark grey) day-ahead forecasts (Tuesday-Saturday) and our model (dark red)



3.5 Combining forecasts

So far we have analyzed our model with the benchmark. We presented each seasonality through error measurements and discussed the results. Since our beginning model is adjusted so it can be forecasted at the same hour, models are closer to the specifications and we would like to see if the overall results are improved by combining them.

3.5.1 Combination using OLS

For the regression method of combining the forecasts, we will use out-of-sample year 2013 for the estimation. Estimation is done by OLS regression and we will compare results with results from previous section. Estimated coefficients of the regression are shown in equation 16.

$$Combined = 5.151082 + 0.5281595 \cdot Our\ model_t + 0.4671016 \cdot Benchmark_t \quad (16)$$

3.5.2 Combination using simple average

For the simple average we do not require an additional out-of-sample year, but since we used year 2013 for regression coefficients estimation, we will report results with and without including year 2013 into error measurements.

Table 8. Error measurement for combined forecasts with and without holiday (period 2013 and onwards)

	With holidays	Without holidays
MAE OLS	33	32
MAE Average	33	32
MSE OLS	1,940	1,756
MSE Average	1,937	1,758
MAPE OLS (in %)	2.349	2.220
MAPE Average (in %)	2.348	2.221
Theil's U-statistic	1.002	1.000

Table 9. Error measurement for combined forecasts with and without holiday (period 2014 and onwards)

	With holidays	Without holidays
MAE OLS	32	32
MAE Average	32	31
MSE OLS	1,873	1,734
MSE Average	1,867	1,732
MAPE OLS (in %)	2.306	2.199
MAPE Average (in %)	2.304	2.198
Theil's U-statistic	1.001	0.999

Results are shown in table 8 and table 9. Combining forecasts by both methods show significantly better results than using individual forecasts. MAPE is lower for almost half a percent by using both methods. A similarly scaled improvement can be seen when holidays excluded.

Results between OLS regression and simple average in this case are very similar. That is expected since in previous analysis models did not show any larger difference in results. If we regressed models in the in-sample period, larger weights for combining forecasts would be assigned to our model. Then results would be in favor of simple average. In this case we find no reason to use regression over simple average since two models which are used in combination have very close results. That is why regression weights are close to average weights.

Much better forecasting results are a confirmation of our error analysis in previous sections. Because most of the average errors from both models have different signs, by combining them, some of the errors cancelled each other. That, in the end, resulted in a much more precise forecasting model. Also, the benchmark model is handling better special events. This in turn led to a decrease in errors in the combined model.

CONCLUSION

In the thesis we presented a model which can be used in practice on a daily basis for forecasting day-ahead electricity demand. Motivation for the model was lack of the publicly available forecasts in practice for this region. In case that they are available, time of the publication can be an issue. We have tried to find what the policies are and the timeline of publishing forecasts in some neighboring countries (and some countries which have their own electricity markets, like Romania) and presented findings.

Theoretical research has been done by many authors. Since statistical models are more used by the economists, we focused on building a statistical model. We analyzed electricity demand and presented its specifics, seasonalities and external factors which have influenced it.

Process of data collecting we explained in data mining section. Also we discussed filters for error detection and methods to deal with errors in data. Weather factors were a second source of data on which we applied similar techniques. In the end data was merged into one time-series and readied for further analysis.

Next step was building the model. We proposed two models with same variable specification. For the estimation period we used three years of data, from 2010 to end of the 2012. For the period of 2013 until the April 2016 by using a minimum available lag structure, we made day-ahead direct forecasts. Results of the forecasts were compared and we chose the better suited one which was also the simpler model. Since electricity demand has three distinctive seasonalities, we compared the models on all three of them. Criteria for choosing the better model were error measurements over all seasonalities: mean error which showed the direction of over- or underestimation, mean absolute error, mean squared error, mean absolute percentage error as well as Theil's U-statistic. For the benchmark we used forecasting data which is published by ELES. Proposed model is adjusted so the minimal lag structure criteria is satisfied and models are from that perspective comparable. Results showed that our model is very similar to the benchmark and when holidays are excluded, our model is slightly better. Also we tested the theory of combining forecasts in order to improve forecast accuracy which proved to be true. We tested two scenarios, simple average between two models and obtaining coefficients by regression. In this case, there was no proof for using regression since simple average produced the same results as regression.

From the perspective of future improvements, there are few things which we will mention. First, as Ramanathan et al. (1997) discuss, by using a 24 hour model with separate equations, there is a possibility to model each hour separately. For the comparison purposes we kept the model specifications the same for all hours, but we believe that if one would model all hours separately by choosing different lags and different variables for different hours, end model forecast precision would improve. Also, we estimated the model once, then we used the same estimators for whole forecasting period. In real life, model would be re-estimated on much shorter time-scale, so forecasting period would be smaller than estimation which was not the case in our example. That also could lead to improvement of the precision. For further research it would be interesting to see if and how much accuracy of forecasting models would be improved if more models with different specifications or estimation procedure would be included. Since we combined only two models, regression in my case was not necessary for estimation of the weights. Adding more models would find regression estimation useful.

The proposed model is easily applicable in many countries, but in order to get good results, research and data mining procedure which we presented should be followed.

REFERENCE LIST

1. Alfares, H. K., & Nazeeruddin, M. (2002). Electric load forecasting: literature survey and classification of methods. *International Journal of Systems Science, 33*(1), 23-34.
2. Berk, K. (2015). *Modeling and Forecasting Electricity Demand: A Risk Management Perspective.* Wiesbaden: Springer.
3. Box, G. E., & Jenkins, G. M. (1976). *Time series analysis: forecasting and control.* San Francisco: Holden-Day.
4. Chatfield, C., & Yar, M. (1988). Holt-Winters forecasting: some practical issues. *The Statistician*, 129-140.
5. Chen, B. J., & Chang, M. W. (2004). Load forecasting using support vector machines: A study on EUNITE competition 2001. *IEEE Transactions on Power Systems, 19*(4), 1821-1830.
6. Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International journal of forecasting, 5*(4), 559-583.
7. Clements, A. E., Hurn, A. S., & Li, Z. (2016). Forecasting day-ahead electricity load using a multiple equation time series approach. *European Journal of Operational Research, 251*(2), 522-530.
8. Darbellay, G. A., & Slama, M. (2000). Forecasting the short-term demand for electricity: Do neural networks stand a better chance? *International Journal of Forecasting, 16*(1), 71-83.
9. Do, L. P., Lin, K. H., & Molnár, P. (2016). Electricity consumption modelling: A case of Germany. *Economic Modelling, 55*, 92-101.
10. Elektromreža Srbije. Retrieved July 10, 2016, from EMS Transparency: http://transparency.ems.rs/tr_load_03.php
11. ELES. *Load and Generation*. Retrieved July 10, 2016, from http://www.eles.si/en/load-and-generation.aspx
12. Engle, R. F., Chowdhury, M., & Rice, J. (1992). Modelling peak electricity demand. *Journal of Forecasting, 11*(3), 241–251.
13. ENTSO-e. Retrieved July 10, 2016, from https://transparency.entsoe.eu/
14. EPEX SPOT. *Intraday market with delivery on the French TSO zone*. Retrieved July 15, 2016, from http://www.epexspot.com/en/product-info/intradaycontinuous
15. Fan, S., & Hyndman, R. J. (2012). Short-term load forecasting based on a semi-parametric additive model. *IEEE Transactions on Power Systems, 27*(1), 134-141.
16. Fay, D., Ringwood, J. V., Condon, M., & Kelly, M. (2003). 24-h electrical load data—a sequential or partitioned time series? *Neurocomputing, 55*(3), 469-498.
17. Feinberg, E. A., & Genethliou, D. (2005). Load forecasting. *Applied mathematics for restructured electric power systems*, 269-285.
18. Government of the Republic of Slovenia. *Holidays*. Retrieved July 24, 2016, from: http://www.vlada.si/en/about_slovenia/political_system/holidays/

19. Hippert, H. S., Pedreira, C. E., & Souza, R. C. (2001). Neural Networks for Short-Term Load Forecasting: A review and Evaluation. *IEEE, 16*(1), 44-55.
20. Holt, C. C. (1957). Forecasting seasonals and trends by exponentially weighted averages, ONR memorandum no. 52, Carnegie Institute of Technology, Pittsburgh, USA published in *International Journal of Forecasting 2004, 20*, 5-10.
21. Hor, C. L., Watson, S. J., & Majithia, S. (2005). Analyzing the impact of weather variables on monthly electricity demand. *IEEE transactions on power systems, 20*(4), 2078-2085.
22. Hyde, O., & Hodnett, P. F. (1997). An adaptable automated procedure for short-term electricity load forecasting. *IEEE Transactions on Power Systems, 12*(1), 84-94.
23. Hyndman, R., & Koehler, A. (2006). Another look at measures of forecast accuracy. *International journal of forecasting, 22*(4), 679-688.
24. Marcellino, M., Stock, J. H., & Watson, M. W. (2006). A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of econometrics , 135*(1), 499-526.
25. MAVIR. *Hungarian Power System actual data: Consumption*. Retrieved July 25, 2016, from: https://www.mavir.hu/web/mavir-en/consumption
26. Ministry of the enviroment and spatial agency - Slovenian Enviroment Agency. Retrieved June 11, 2016, from: http://meteo.arso.gov.si/met/en/app/webmet/
27. Nowotarski, J., Liu, B., Weron, R., & Hong, T. (2016). Improving short term load forecast accuracy via combining sister forecasts. *Energy, 98*, 40-49.
28. OKTE. *Total STM results CZ – SK – HU – RO*. Retrieved July 25, 2016, from: https://www.okte.sk/en/short-term-market/published-information/total-stm-results-cz-sk-hu-ro
29. Pardo, A., Meneu, V., & Valor, E. (2002). Temperature and seasonality influences on Spanish electricity load. *Energy Economics, 24*(1), 55-70.
30. Ramanathan, R., Engle, R., Granger, C. W., Vahid-Araghi, F., & Brace, C. (1997, June). Short-run forecasts of electricity loads and peaks. *International Journal of Forecasting, 13*(2), 161-174.
31. Ružić, S., Vučković, A., & Nikolić, N. (2003). Weather Sensitive Method for Short Term Load Forecasting in Electric Power Utility of Serbia. *IEEE Transactions on Power Systems, 18*(4), 1581-1586.
32. Sensfuss, F., Ragwitz, M., & Genoese, M. (2008). The merit-order effect: A detailed analysis of the price effect of renewable electricity generation on spot market prices in Germany. *Energy policy, 36*(8), 3086-3094.
33. Suganthi, L., & Samuel, A. A. (2011). Energy models for demand forecasting—A review. *Renewable and sustainable energy reviews, XVI*(2), 1223-1240.
34. Taieb, S. B., & Hyndman, R. J. (2012). Recursive and direct multi-step forecasting: the best of both worlds.
35. Taylor, J. W. (2003). Short-term electricity demand forecasting using double seasonal exponential smoothing. *Journal of the Operational Research Society, 54*(8), 799-805.

36. Taylor, J. W., De Menezes, L. M., & McSharry, P. E. (2006). A comparison of univariate methods for forecasting electricity demand up to a day ahead. *International Journal of Forecasting, 22*(1), 1-16.
37. Tofallis, C. (2015). A Better Measure of Relative Prediction Accuracy for Model Selection and Model Estimation. *Journal of the Operational Research Society, 66*, 1352-1362.
38. Transelectrica. *Data for electricity demand, production, exchange*. Retrieved July 24, 2016, from: http://transelectrica.ro/widget/web/tel/sen-grafic/-/SENGrafic_WAR_SENGraficportlet?display=APE
39. Transelectrica. *Required Transparency of System Load*. Retrieved July 11, 2016, from http://www.transelectrica.ro/en/web/tel/consum
40. Valor, E., Meneu, V., & Caselles, V. (2001). Daily air temperature and electricity load in Spain. *Journal of applied Meteorology, 40*(8), 1413-1421.
41. Wang, P., Liu, B., & Hong, T. (2016). Electric load forecasting with recency effect: A big data approach. *International Journal of Forecasting, 32*(3), 585-597.
42. Weron, R. (2006). *Modeling and forecasting electricity loads and prices: a statistical approach.* Chichester, West Sussex, England: John Wiley & Sons Ltd.
43. Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting, 30*, 1030-1081.
44. Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science, 6*(3), 324-342.