UNIVERSITY OF LJUBLJANA
FACULTY OF ECONOMICS

MASTER'S THESIS

HUNGARIAN POWER EXCHANGE (HUPX) SPOT PRICE ANALYSIS

Ljubljana, September 2014 MARKO HALUŽAN

AUTHORSHIP STATEMENT

The undersigned Marko Halužan, a student at the University of Ljubljana, Faculty of Economics, (hereafter: FELU), declare that I am the author of the master's thesis entitled Hungarian power exchange (HUPX) spot price analysis, written under supervision of prof. dr. Miroslav Verbič.

In accordance with the Copyright and Related Rights Act (Official Gazette of the Republic of Slovenia, Nr. 21/1995 with changes and amendments) I allow the text of my master's thesis to be published on the FELU website.

I further declare

- the text of my master's thesis to be based on the results of my own research;
- the text of my bachelor thesis / master's thesis 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 my bachelor thesis / master's thesis / doctoral dissertation in accordance with the FELU's Technical Guidelines for Written Works and
    - obtained (and referred to in my bachelor thesis master's thesis) all the necessary permits to use the works of other authors which are entirely (in written or graphical form) used in my text;
- 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 (Official Gazette of the Republic of Slovenia, Nr. 55/2008 with changes and amendments);
- to be aware of the consequences a proven plagiarism charge based on the submitted bachelor thesis / master's thesis could have for my status at the FELU in accordance with the relevant FELU Rules on Master's Thesis

Ljubljana, 2.9.2014 Author's signature

TABEL OF CONTENTS

TABEL OF FIGURES

TABEL OF TABLES

INTRODUCTION

There are many Power Exchanges (hereinafter: PXs) in Europe. A number of papers deal with price analysis and price forecasts on these exchanges. The Hungarian Power Exchange (hereinafter: HUPX) is one of the least explored exchanges as some of the major price forecasting providers have only just recently added HUPX price prediction to their service. Hence the HUPX spot price analysis is the subject of my study.

The products that can be traded on the Hungarian Power Exchange day-ahead market (hereinafter: HUPXDAM) are standard hourly contracts for the day-ahead physical delivery of electricity within the Hungarian power grid (HUPX, 2014c). These contracts are sold at the price €/MWh. In day-ahead markets, separate prices are quoted for delivery in each specific hour in the next day; the daily average is then the average over the 24 hours (Huisman, Huurman, & Mahieu, 2007). The daily average price is commonly known as the Base Price.

HUPX can be considered as a balancing point due to its strategic position. There are only two PXs (BSP SouthPool & Opcom) in the region that could be considered as a substitute for HUPX, but they are not as convenient as HUPX. HUPX represents a major balancing point for the broader Balkan region (Bosnia, Croatia, Serbia, Slovenia, Romania), which means that energy surpluses and shortages in these countries are balanced at HUPX. Any imbalance between supply and demand causes the system frequency to deviate from standard (Stoft, 2002). To this end, the supply and demand in a particular power grid always have to be balanced and that is a specific feature common to electricity markets. This feature separates electricity from other commodities. The facts that electricity has low demand elasticity and HUPX is a balancing point for the wider region, both point to an unpredictable price behaviour and high volatility.

Market risk related to trading is considerable due to extreme volatility of electricity prices. This is especially true for spot prices, where volatility can be as high as 50% on the daily scale, i.e. over ten times higher than for other energy products (natural gas and crude oil) (Misiorek & Weron, 2005). For placing a reasonable bid at Power Exchange, it is crucial to have a price forecast or at least a good understanding of the main spot price drivers.

Considering HUPX as a balancing point of the Balkans, I will examine the impact of outside temperature, hydro and wind production in the Balkan region on the HUPXDAM clearing price for Base (average of 24 hours). In my thesis I will use an ARMA price forecasting model, however, it will be extended with exogenous variables (daily average temperature, daily average wind production, daily average river stream) into the ARMAX model. Hence the ARMAX model will serve for forecasting the HUPX spot price and examining the exogenous effects on the HUPX spot price. The outcome of the thesis will result in a better understanding of the HUPX price behaviour.

First hypothesis: Daily average temperature in the Balkans is a relevant and statistically significant determinant of the HUPX market clearing price forecast.

Second hypothesis: Daily average wind production in the Balkans is a relevant and statistically significant determinant of the HUPX market clearing price forecast.

Third hypothesis: Daily average hydro production in the Balkans is a relevant and statistically significant determinant of the the HUPX market clearing price forecast.

1 POWER MARKET

In recent years, all European countries have moved from regulated regional monopolies to liberalised electricity markets (Ockenfels, Grimm, & Zoettl, 2008). Under regulation, in fact, price variation was minimal and under the strict control of public-owned commissions, which determined tariffs on the basis of average production costs. In this controlled environment the attention was focused on demand forecasting. In particular, the most sophisticated statistical techniques have been proposed to achieve satisfactory short-run predictions. On the other hand, under deregulation (liberalization), price formation was delegated to the law of supply and demand. Because of the distinct characteristics of electricity, price volatility in liberalised markets has increased far beyond those of any other commodity or financial asset. Therefore, great interest has been placed on developing accurate price forecasting models (Fezzi, 2007).

The power market's liberalisation has led to a change from a centralised structure, where the only variable of interest in terms of prediction was demand, to a competitive environment, where the prediction of price is an unavoidable task for both producers and consumers (García-Martos, Rodríguez, & Sánchez, 2012).

The main characteristics of electricity are: non-storability; production and consumption have to be balanced; essential and homogenous commodity; physical and contract flows are different; low demand elasticity. To deliver electricity to a certain point (i.e. from country A to country B), cross-border capacities must be ensured. In most cases, the flow goes from the region with a low price (energy surpluses) to the region with a higher price (energy shortages). This is usually a consequence of either having more energy from a cheaper source or a lower demand in one region, compared to another region. Price differences can be constant because of the composition of energy generating sources. They can also be just a result of instant higher energy availability from the sources with lower marginal costs of production.

Inelastic demand is a characteristic of goods that are a necessity. The substitutes for these goods are few and difficult to obtain. Power and electricity fall into this category as they are vital goods seen as essential for the existence of modern civilisation (EWEA, 2010). As demand is inelastic, minor changes can result in major price changes.

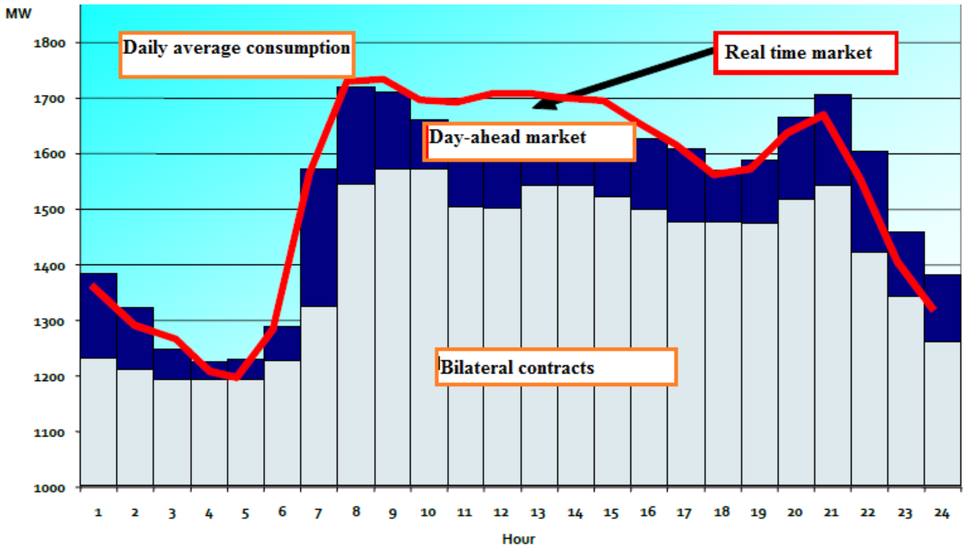
In market terms, suppliers are supplying energy to final consumers representing their aggregated demand because small consumers do not have direct access to the market. In turn, producers are representing the aggregate supply curve in the electricity market.

Harris (2006) describes the demand profile using the following key pieces of information:

1. Last year's and previous year's profile.
2. Demand trend based on gross domestic product (GDP) growth and trends in energy intensity.
3. Weather forecast, particularly temperature, wind chill and cloud cover, and trends such as global warming and urbanisation. Also special events such as hurricanes and solar eclipse.
4. Time of dawn and dusk.
5. Bank holidays, television schedules and other consumer diary events.
6. Trends in domestic equipment such as air conditioners and equipment with clocks.
7. Changes in financial incentives to alter consumption such as off peak rates supported by meter clocks.
8. Installation, trends, prevailing conditions and other factors for embedded generation.
9. Changes to transmission and distribution infrastructure, particularly constraints and losses.
10. Economic incentives for reducing losses and constraints, and the impact on participant behaviour.

Electricity demand can be forecasted using the above stated information. The electricity market is usually segmented into the following categories: Residential, Industrial, Commercial, Transportation, Other. In Figure 1 we can see how the diagram of daily consumption of electricity is covered. In general, the demand for electricity is influenced by some social and economic activities and by weather conditions (Liu & Shi, 2013).

Figure 1. The Diagram of Daily Consumption of Electricity



Source: Gubina A., *Electricity as a commodity*, 2010.

The majority of the diagram is covered in advance on the forward market with bilateral contracts. The day-ahead market covers a small amount of consumption and the real time market is even smaller. The real time market in most cases serves for purchasing replacement power in case of aggregate failure or for balancing the production of renewable sources (i.e. wind and photovoltaic production).

The demand curve is constructed from aggregated demand bids, which is why the bidding strategy of customers has a significant impact on the position of the intersection point, i.e. the final values of the market clearing price and the market clearing volume on the spot market. Natural seasonal load variation of the system has, of course, an impact on the level of purchased electricity (demand) through spot markets, which causes natural seasonal price variations as well (Kolcun, Oleinkova, & Truicki, 2012).

Figure 1 will also help to define basic market products and their characteristics, according to Gubina (2010):

1. BASE (00–24)
   a) Constant consumption, affordable price.
2. EU PEAK (08–20)
   a) More expensive than Base.
   b) Higher consumption.
   c) Buyer pays a premium for the option to consume only a part of the day.
3. EU OFF PEAK (00–08 and 20–00)
   a) Consumption is low during the night.
   a) Energy surpluses (due to start-up costs).

b) Low price.

Energy can be provided by different power generators. Generators are using different technologies for generating electricity; hence they deal with different costs of production. According to Murray (2009), the costs of generation that need to be recovered from the sale of energy through the life of the plant will include:

1. The capital costs of the plant and interest incurred during construction expressed as an annuity or annual charge.
2. The cost of fuel used in the production of energy that is exported and used internally by auxiliaries.
3. The fixed operating costs, e.g. staff, insurance and transport, which do not vary with the plant utilisation.
4. The variable operating costs, e.g. maintenance material and labour costs, which will be influenced by the plant utilisation and wear and tear.

Expressed in the basic form, the total costs of electricity produced in the power plant consist of fixed and variable costs as in equation (1):

$$C_{total} = C_{fix} + C_{var} \tag{1}$$

Fixed costs C_{fix} are related mostly to investment and economic profit to be earned, in short-run its operating and maintenance costs, wages, depreciation, social fund and other obligations that are fixed and independent of the level of production. Variable costs C_{var} are all costs that depend on the level of production in a particular power plant. They mostly cover the costs of fuel and emission allowances. However, once a plant is commissioned, the marginal cost of producing an additional unit of electricity should determine its operation (dispatch). Marginal costs roughly correspond to fuel costs and costs to purchase emission allowances as their volume also depends on the level of production. This fact is important because it determines the merit order of power plants in the supply curve according to marginal costs, thereby the marginal generator and thus market clearing price on the electricity market (Kolcun et al., 2012).

Power plants fired with fossil fuels (e.g. brown coal, hard coal, gas, oil) have additional costs with buying CO_2 coupons for CO_2 emissions caused by their operation. Concerns with global warming have accelerated, with the power sector as a focus of attention for reducing emissions. In practice this means burning less coal, gas and coal and substituting them with renewable and alternative energy sources with lower emission levels (Murray, 2009).

Reducing emissions is promoted by special incentives schemes for renewable sources (the majority for photovoltaic and wind production). The incentives schemes are necessary because the private sector would not invest in renewables, since investing in these power plants that are based on market prices is not lucrative. On the other hand, these power plants are usually small and their unpredictable production renders them inappropriate to operate alone in market competition.

These schemes are designed to provide an incentive for the development of renewable sources by providing additional revenue over and above that derived from the sale of the energy. They may take the form of a “feed-in” tariff whereby there is a guaranteed income price for a number of years. These schemes do not generally take account of the varying market energy cost but fix the rate for the contract term (Murray, 2009).

Trading for the power delivered in any particular minute begins years in advance and continues until real time, the actual time at which the power flows out of a generator and into a load. This is accomplished by a sequence of overlapping markets, the earliest of which are forward markets that trade nonstandard, long-term forward contracts. Futures contracts are standardized, exchange-traded, forward contracts. Most informal forward trading stops about one day prior to real time. At that point, the system operator holds its day-ahead market. This is often followed by an hour-ahead market and real time market (Stoft, 2002). In the present thesis, the spot market will refer to the day-ahead market (DAM).

Contracts for larger blocks of energy making up a large proportion of the requirement for the period ahead may be established through a tendering process (Murray, 2009). Bilateral trading stands for direct deals between producers and suppliers; these deals can be referred to also as over-the-counter (hereinafter; OTC) trading.

Fine tuning of positions is in most cases done via power exchanges. Power exchange is convenient for fine tuning, since single hours can be bought or sold. Fine tuning is in most cases done day-ahead prior physical delivery, because market players have up-to-date information on consumption and supply. Power exchanges have been established to operate in a similar manner to stock exchanges. They use IT systems to display current bid and offer volumes and prices, and enable clients to establish deals for physical delivery on an anonymous basis (Murray, 2009).

On the European energy market, a large part of the energy is traded in long-term contracts, while only a comparatively small part is traded day-ahead in the spot market auctions (Ockenfels et al., 2008). The time line of trading for a specific delivery day would show that the trading volume is decreasing in time. This is a consequence of risk management.

The suppliers will have to agree contract prices with consumers and set tariff prices for small users in advance, thus being exposed to a risk of prices for wholesale energy being higher. The risk is minimised by contracting ahead for energy from generators for the period for which the costumer prices are fixed. The other source of risk to the supplier is the volume risk resulting from not having an accurate estimate of the final demand of the customer base (Murray, 2009).

The generators are also exposed to price risk and have to decide what and when to contract for their expected output. They have to take similar decisions in how far ahead to contract depending on whether they expect prices to rise or fall. They are also exposed to balancing market prices as they may have plant problems that prevent them from meeting their contracted commitment. Generators also have to manage the volume risk and decide on how much output would be concentrated in the long term as opposed to trading on the spot market at the day-ahead stage or bidding into a balancing market, where prices may be higher. The other risk in the short term is predicting the number of hours of operation over which start-up costs have to be recovered (Murray, 2009).

Electricity has been traded across national borders increasingly since the end of the 1990s. The flows of electricity here primarily follow the different price levels in Europe. However, the transmission capacities at the grid connectors between the individual European neighbours are limited. The consequence is that today, the demand for transmission capacities for cross-border trading exceeds the available capacity of the lines at many locations, resulting in cross-border bottlenecks. If all requests were permitted, then cable overloads would inevitably be the consequence (Tennet, 2014).

Congestion management resolves border bottlenecks by auctioning capacities (annual, monthly and daily auctions). A daily capacity auction is carried for every single hour of the day for day-ahead.

1.1 Spot (day-ahead) market price determination

Many power markets rely on a central day-ahead auction in which generators submit individual supply curves and the system operator uses these to determine the market price. The competitive price is determined by the intersection of the market's supply and demand curve. Marginal costs only determine the supply curve. Day-ahead markets run by system operators take the form of either exchanges or pools and are operated as auctions. The process of selecting the winning bids is often complicated by transmission and generation constraints which can require the use of enormously complex calculations and sophisticated mathematics (Stoft, 2002). Producers submit to the Market Operator (MO) production bids that typically consist of a set of energy blocks and their corresponding minimum selling prices for every hour of the market horizon. Analogously, retailers and large consumers submit to the MO consumption bids that consist of a set of energy blocks and their corresponding maximum buying prices. The MO uses a market-clearing algorithm to clear the market, which results in a market-clearing price as well as the scheduled production and consumption for every hour of the market horizon. The market-clearing price is the price to be paid by retailers and to be charged by producers (Conejo, Contreras, Espínola, & Plazas, 2005).

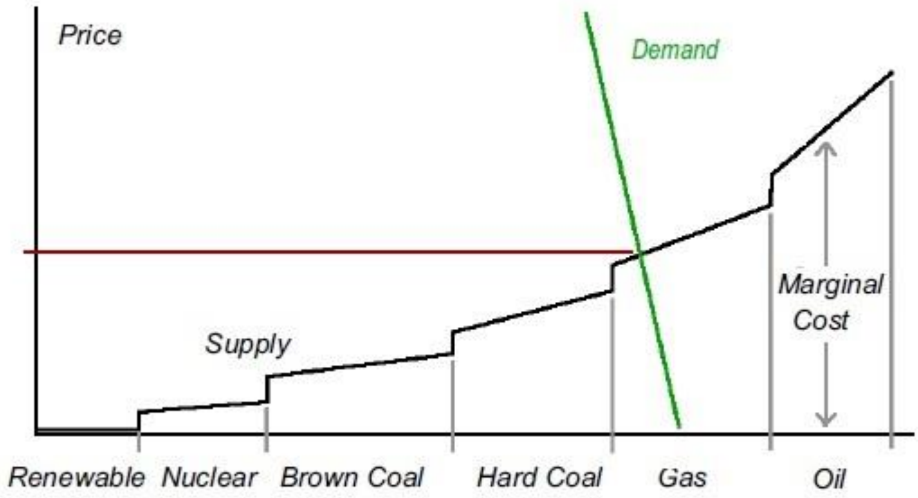
On the day-ahead market, standard hourly contracts are sold for the day-ahead physical delivery. Some electronic markets also allow submitting block bids consisting of random blocks of hours. Trading is done on an anonymous basis, and accepted bids are settled at the determined corresponding hourly price. The availability of renewable sources generation has a major impact on the spot price determination because their production depends on the weather (unpredictable in advance). Twenty-four hours prior to the physical delivery, renewable source generator owners have at their disposal reliable weather forecasts and can place their bid on the PX accordingly.

Electricity can be generated from different sources. To understand price determination at the spot market, it is important to be familiar with the merit order of power plants by their short run marginal costs of production. Marginal costs of production eventually become the costs of producing an additional unit of output. For day-ahead auctions, this is reasonable since the majority of aggregates have already sold a part of their production on the forward market and will be in function on the delivery day (no start-up costs etc.).

Marginal costs play a key role in the economic theory that proves that a competitive market is efficient, but there are also two practical uses of marginal costs that increase its importance in a power market. First, many power markets rely on a central day-ahead auction in which generators submit individual supply curves and the system operator uses these to determine the market price. Because price should equal marginal cost in an efficient market, the auction rules should be informed by a coherent theory of marginal costs. Second, many power markets suffer from potential market-power problems which cause the market price to diverge from marginal costs. Market monitors need to understand this divergence (Stoft, 2002).

In figure 2 we can see that renewable sources (e.g. hydro, photovoltaic, wind power plants) are first supplying energy to the market because their marginal costs are negligible. It would be reasonable that every power plant places bids at the spot market according to its marginal costs. No supplier is ready to sell a unit of electricity at a price which is lower than the additional costs of this unit (Ockenfels et al., 2008). The last power plant supplying energy is the one that sets the market clearing price (gas power station in Figure 2).

Figure 2. Merit Order Pricing



Source: Cabrera, B. L., & Schulz F., *Probabilistic forecasts of electricity spot prices using residual load,* 2014, p.22.

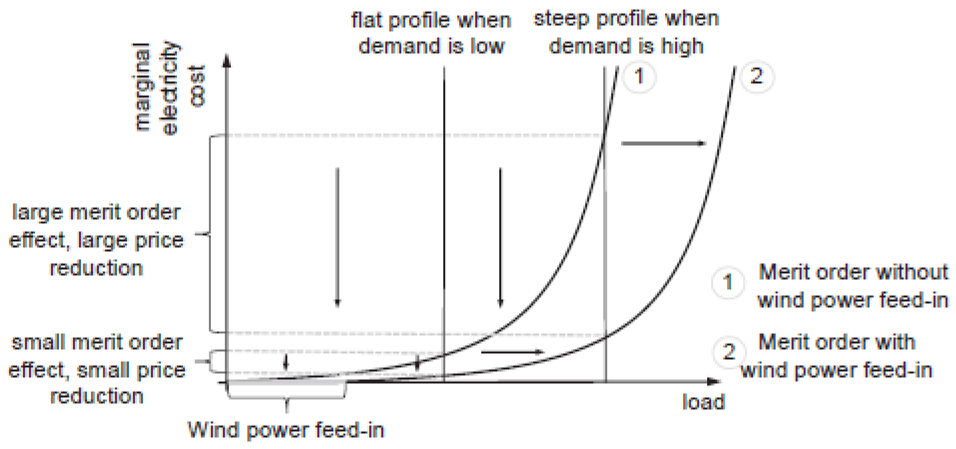
Adding wind power into the generation mix will affect the supply curve, the supply curve will shift, and a new price will be determined as a result of market dynamics (EWEA, 2010). Increased hydro[1] or photovoltaic power generation also causes major shifts in the aggregated supply curve because they are like wind, the sources with the lowest marginal cost. This means that these sources are the first to serve energy when they are available and will always affect market dynamics. This phenomenon is called the Merit Order Effect (Genoese, Ragwitz, & Sensfuß, 2007).

Supporting renewables to achieve a less polluting and (foreign) dependent energy sector has many consequences. Among them, an increased renewable production of electricity crowds out other high(er) marginal-cost technologies and results in lower electricity prices (Würzburg, Labandeira, & Linares, 2013). These lower prices basically come from the fact that renewables bid into wholesale electricity markets at almost-zero prices and therefore shift the electricity supply curve to the right (Würzburg et al., 2013).

Figure 3 shows the crowding out effect. Due to the wind power feed-in, the aggregated supply curve shifts to the right, causing price reduction. Price reduction occurs because aggregates with a higher marginal cost are crowded out. Due to the shape of the supply and demand curve, a price reduction is much higher when demand is high compared to when it is low. The shape of the supply curve is determined by short-run marginal costs, while the inelastic demand is represented as a straight line.

[1] Hydro turbines convert the potential energy of water into electricity by using water pressure to drive pumps. The water sources are lakes (filled naturaly or by pumps) and rivers (Harris, 2006). Water sources filled by pumps are not considered in this case due to pumping costs.

Figure 3. Right Shift of the Merit Order and the Supply Curve, particularly due to Wind Power Feed-in



Source: Keles, D., Genoese, M., Möst, D., Ortlieb, S., & Fichtner, W., *A combined modeling approach for wind power feed-in and electricity spot prices*, 2013, p. 214.

Figure 1 indicates that the majority of demand is covered with forward contracts. Forward contracts are tending to cover a predicted average demand for a specific period (e.g. week, month, quarter, a year), while spot market serves for fine tuning. Every deviation from the predicted average demand that could not have been predicted in advance is covered at the spot market. These deviations are mainly caused by weather and unpredictable events. Weather has an impact on the usage of heating and air conditioning appliances. Outside temperature values above a historical average cause a major shift in power demand on the spot market.

Unpredictable events influencing the market are unpredicted generator outages and technical issues with cross-border transmission infrastructure. Electricity demand and supply have to be balanced; as a result, such issues are reflected on the spot market. A generator outage has to be replaced with energy from another generator, which in most cases figures higher on the merit order scale.

Energy flows are going from a region with energy surpluses (a region with a low price) to a region with energy deficits (a region with a high price). Technical issues with cross-border transmission infrastructure have an impact on the demand side and also on the supply side because foreign supply or demand cannot physically reach the market. If a market has a power deficit and cheaper energy cannot be imported due to technical issues with the cross-border capacities, the market becomes isolated and aggregates with higher marginal costs are supplying power to the market. In the opposite case, a market has energy surpluses and these surpluses cannot be distributed to the market with an energy deficit.

In practice, the daily capacity auction price has a great impact on the spot and hence sends a signal to traders and generator owners. Daily capacity auctions are held before PX auctions. Consequently, the results of capacity auctions are already reflecting the energy situation in a broader region as well as the availability of cross-border capacities. Power traders want to refund capacity costs, which is why these costs are taken into account in their bidding strategies.

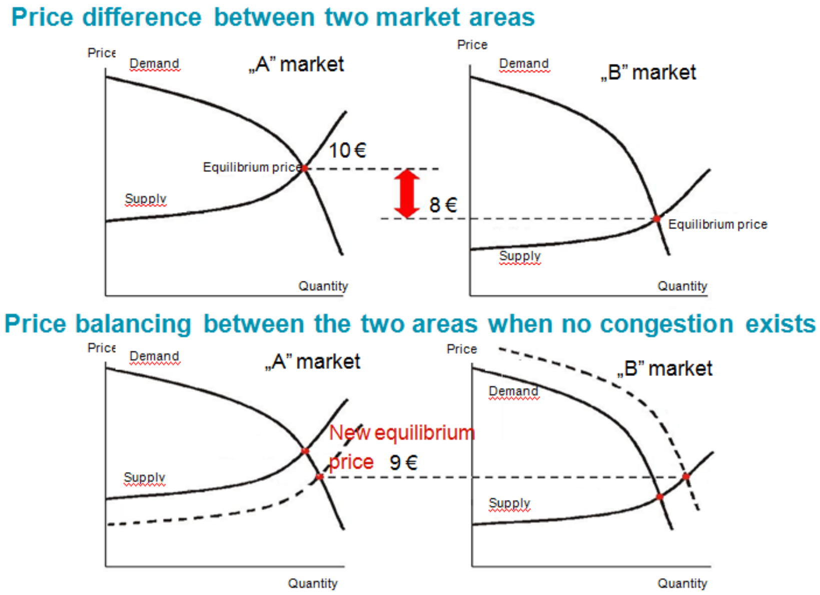
1.2 Hupx spot price

As part of the liberalization of the Hungarian energy sector, the Organized Electricity Market was launched in July 2010 as a subsidiary of MAVIR ZRt (Hungarian Transmission System Operator). Since then, the Hungarian power exchange (HUPX) has built a reputation of being an inevitable platform of domestic power trading (Epexspot, 2014a).

Since 12 September 2012, HUPX has been coupled with OTE (Czech) and OKTE (Slovakia) power exchanges. Market coupling denotes cross-border matching of energy supply and demand while taking into account the available cross-border capacity provided by the Transmission System Operators. All inputs (exchange bids and capacity data) are considered together. The method is almost the same as local matching of bids, but uses cross-border capacity profiles as constraints of trade between market areas. The energy then flows from low price to high price areas in order to balance the market prices. In case of no congestion (ample capacity), the prices become equal in the concerned market areas. In case of congestion (scarce capacity), one price area has a lower price and the second one a higher one. The price difference (spread) between neighbouring market areas' prices represents the capacity auction price (HUPX, 2014b).

In Figure 4 we can see the principle of market coupling. In case of no market coupling, the formation of two different prices is visible. The spread between the areas is € 2. In case of no congestion issues (enough capacities between the areas), we can see the formation of one single price at € 9. On the other hand, when there is a congestion issue, two different prices are formed.

Figure 4. Market Coupling Principle



Source: HUPX, *Market coupling*, 2014b.

For price analysis, it is crucial to understand the power situation in the broader region. It is important to obtain an overview of power plant units in the region and how they can reach HUPX (cross-border transmission capacities). Table 1 shows that Hungary has a good grid connection with all its neighbouring countries, except with Slovenia (the transmission line is under construction). It has the best connection with the Slovak Republic and Ukraine. The Balkan region can be accessed on all 3 borders (Croatia, Serbia and Romania).

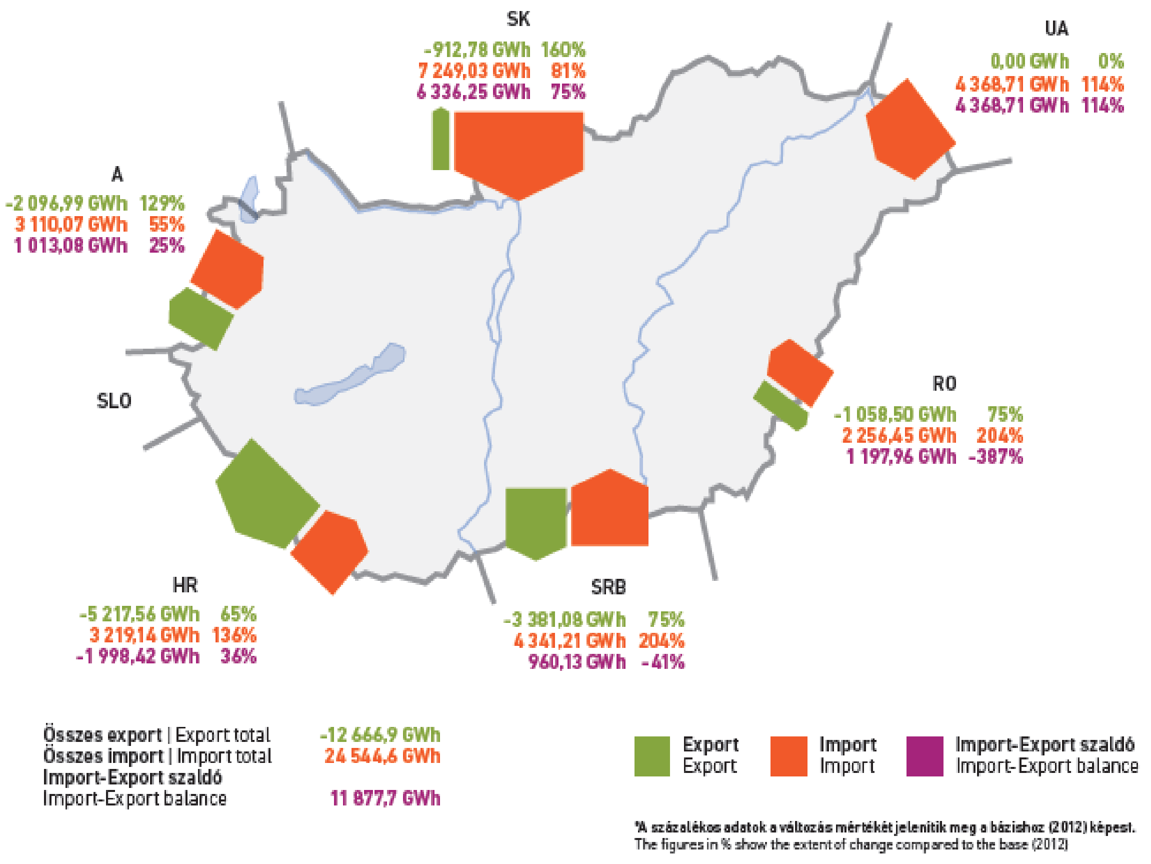
Table 1. Cross-border Interconnections

Country	Transmission line	Net transfer capacity (from/to)
Slovak Republic	2x400 kV	1250/800 MW
Ukraine	1x750 kV, 1x400 kV, 2x220 kV	450-1100 MW
Romania	2x400 kV	600 MW
Serbia	1x400 kV	600 MW
Croatia	2x400 kV (double circuit each)	600/1000 MW
Austria	2x400 kV, 2x220 kV	500 MW
Slovenia	Under construction	

Source: IEA, *Energy Policies of IEA Countries, Hungary 2011 Review*, 2011, p. 106.

In Figure 5 we can see that electricity transit mainly goes from north to south. Since Hungary lies between Central-Eastern and South-Eastern regions of the European electricity system, maintaining and expanding the interconnection capacity is a long-term concern for both Hungary and the entire region. A project under consideration would add one more 400 kV cross-border connection with the Slovak Republic towards 2020 to strengthen north-south flow capability (IEA, 2011). This new cross-border connection with Slovakia would secure additional capacities for the market coupling. Transit from north to south is reasonable because of the operating market coupling with OTE and OKTE. HUPX is coupled on the Slovak border because electricity prices in Slovakia and the Czech Republic are lower and the flow in this direction is reasonable (as illustrated in Figure 9). Import from Ukraine is reasonable due to vast cross-border capacities and because Ukraine has 15 operating nuclear reactors. The government plans to maintain nuclear share in electricity production to 2030, which will involve substantial new build (WNA, 2014). Nuclear share in Ukraine was 48% of total electricity production in 2009. Nuclear power plants are one of the cheapest energy sources.

Figure 5. International Electricity Exchange 2013



Source: Mavir, *Data of the Hungarian electricity system 2013*, 2013, p.17.

Flows on the Balkan borders are almost equally weighted, with the exception of the border with Romania. A small proportion of export and import on the Romanian border is a result of import-export fees. Romania has a vast wind production capacity. According to EWEA (2013), the installed wind production capacity by the end of 2013 was 2599 MW. A part of this production can be balanced on the Greek or Turkish power market since more attractive prices can be obtained compared to the HUPX spot price. In my opinion, HUPX is more convenient because of the direct grid connection and because Greece and Turkey have a high wind production capacity of their own, which can be to some extent correlated with the wind production in Romania.

Serbian and Croatian borders have fairly similar characteristics. This is partly due to their common history and land configuration as electricity production is based on river-run hydro power plants and lignite-fired thermo units. We can notice that there is more export on the Croatian border compared to the Serbian border. This is mainly a result of the existence of PX in Slovenia, but without a direct grid connection with Hungary. To this end, traders use the route via Croatia to enter Slovenia since it is cheaper than entering Slovenia via Austria and vice versa. An increased consumption profile in Croatia in summer (Figure 14) is also a consequence of higher export towards Croatia. By the end of 2013, 302 MW of wind production facilities were installed in Croatia (EWEA, 2013). Slovenian producers are also balancing a part of their hydro and thermo production on HUPX due to attractive prices. According to Bojnec and Papler (2012), electricity supply in Slovenia from water resources depends on natural conditions, but it is particularly biased on the adverse weather conditions. Electricity production from BiH is also balanced on the borders with Croatia and Serbia. Their production also consists of a mix of hydro and thermo production. All these countries are quite significantly hydro dependent, so when river streams are high, surpluses are balanced at HUPX and vice versa.

Import to Austria is mostly based on low night hours prices on the HUPX spot market (Figure 8) since a lot of lignite or coal-fired power plants in the Balkans are balancing their surpluses on HUPX (due to start-up costs). A higher flow in the other direction is logical due to constant spread in all the remaining hours. The price difference between HUPX and neighbouring PXs can be seen in Figure 8.

Cross-border capacities are allocated through explicit auctions. MAVIR conducts yearly, monthly and daily capacity auctions in co-operation with neighbouring TSOs. The day-ahead capacities on the Hungarian-Slovak border are allocated through implicit auctions. An implicit auction is the basic method of the Czech-Slovak-Hungarian market coupling. Capacity allocation is based on simultaneous consideration of power flows and the available cross-border capacity within the market coupling calculation algorithms (HUPX, 2014b).

One of the concerns in Hungary is its ageing infrastructure and the need to replace network assets. The 750 kV substation Albertirsa in Hungary, connected to the line from Ukraine, is reaching the end of its expected lifetime around 2012. However, any plans to replace it are still under consideration, mainly because it is unclear whether the investment for importing from Ukraine would be viable (IEA, 2011). The 750 kV line to Ukraine remains operative today.

HUPX can be considered as a balancing point of the Balkans owing to its strategic position. There are only two other Power Exchanges (BSP-SouthPool and Opcom) in the region that could be considered as substitutes for HUPX, but they are not that convenient. Figure 5 indicates that HUPX represents a balancing point for the broader Balkan region (Bosnia, Croatia, Serbia, Slovenia) since it is the only convenient PX for balancing in the region. This means that energy surpluses and shortages in these countries are balanced at HUPX.

BSP SouthPool is a Power Exchange in Slovenia and it is coupled with GME in Italy. Market coupling on the Slovenian-Italian border is a joint project involving power exchanges (GME and BSP), a power market operator (Borzen), and transmission system operators (TERNA and ELES) (BSP-SouthPool, 2013). The Slovenian grid is connected to the Balkans only on the Croatian border; the demand in Slovenia is relatively low compared to the other countries, although it results from market coupling reinforced by the demand from Italy.

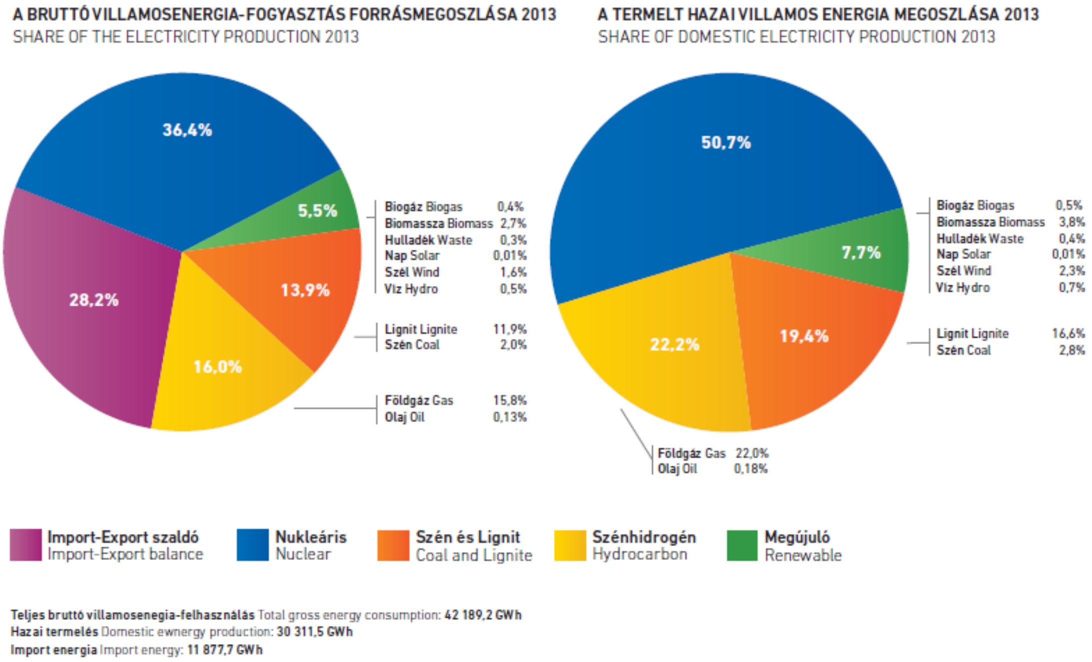
Romania has been operating one of the few day-ahead trading platforms in the region, OPCOM, which gave the country a leading role as the provider of price signals and balancing opportunities for traders in the region. Statkraft has been promoting the further development of the Romanian market, for example through the disposal of the import and export fees or through opening the OPCOM day-ahead market to foreign companies (Statkraft, 2014). OPCOM (Power exchange in Romania) has a good grid connection with the Balkans, but is not as appropriate as HUPX for balancing because of its fees.

In Figure 6 we can see that to keep the power system in balance, Hungary has to import energy. In 2013, 28% of the consumed energy was imported in order to cover gross domestic consumption. A total of 50% of domestic electricity was produced by nuclear power plants. Figure 2 thus shows that this source of energy is relatively cheap.

Hydrocarbon production (gas and oil) covered 22% of the domestic production. Gas and oil are the most expensive energy sources on the merit order scale. Coal and lignite covered additional 20% of domestic production; they are in the middle of the merit order scale by their short-run marginal costs. It is necessary to bear in mind that hydrocarbon, coal and lignite production (almost 40% of the domestic production) results in additional costs through buying CO_2 coupons for their emissions. Nevertheless, energy use in

Hungary produces relatively few CO_2 emissions. In 2009, emissions per GDP were 15% lower than the International Energy Agency (hereinafter; IEA) average. Hungary's low CO_2 emissions are linked to small shares of coal and oil energy supply. Natural gas accounted for 43% of all energy-related CO_2 emissions in 2009, the second largest among IEA member countries (IEA, 2011). The combustion of natural gas produces far less CO_2 emissions than coal and oil combustion.

Figure 6. Composition of Consumption and Production 2013 (in percentage)



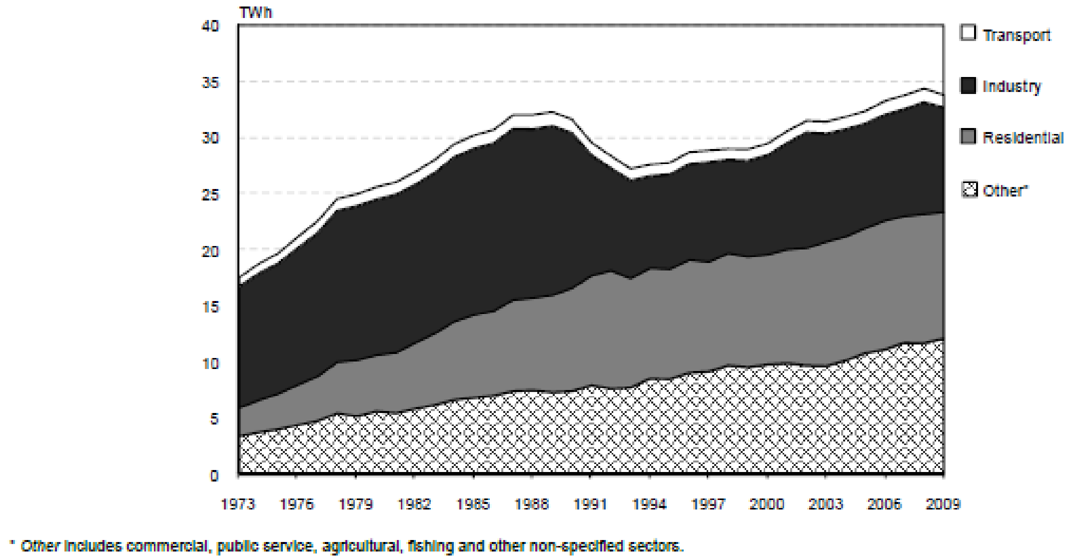
Source: Mavir, *Data of the Hungarian electricity system 2013*, 2013, p.7.

Renewables represent roughly 8% of the domestic production. On the merit order scale, renewable energy sources have the lowest short-run marginal cost. The following renewable sources can be found in Hungary: biogas, biomass, waste, solar, wind and hydro. The only unpredictable production is the weather-dependent production. To this end, their increased production would have a direct impact on the HUPX spot price. The values of the total production from hydro and solar power plants are negligible. Wind production represents 2.3% of total production; we will look into wind production in the following spot price analysis. Hungary grants feed-in tariffs for electricity generation from cogeneration plants and also from renewable energy sources. MAVIR has an obligation to purchase all electricity generated under the feed-in tariff system at a price specified by law (IEA, 2011). By the end of 2013, a wind production capacity of 329 MW was installed in Hungary (EWEA, 2013).

Electricity demand has been steadily increasing from the mid-1990s, growing on average by 1.6% a year over the last decade. In 2009, electricity demand dropped due to the latest global crisis, but still remained higher than in 2007. The government expects electricity consumption per capita to further increase until 2020 and beyond.

Electricity demand typically peaks in winter, however, the summer peak is approaching the winter peak level, mainly on account of the increasing use of air conditioning (IEA, 2011). In Figure 7 we can see that the majority of consumption is consumed by others (i.e. commercial, public service, agricultural, fishing and other non-specified sectors). The residential sector follows with roughly one third of consumption. Most of the electricity consumption is covered by others and the residential sector. The consumption of these two sectors mostly depends on temperature variations from a certain historical average. These temperature variations are mainly covered on the spot market since they are hard to predict in advance. Industrial and transport consumption is more or less stable over time. Orders for their services are usually fixed over time or are easy to predict (economic activity, GDP etc.); hence the variation of their consumption related to their core activity is negligible and not that relevant for the spot market.

Figure 7. Electricity Consumption by Sector, 1973 to 2009



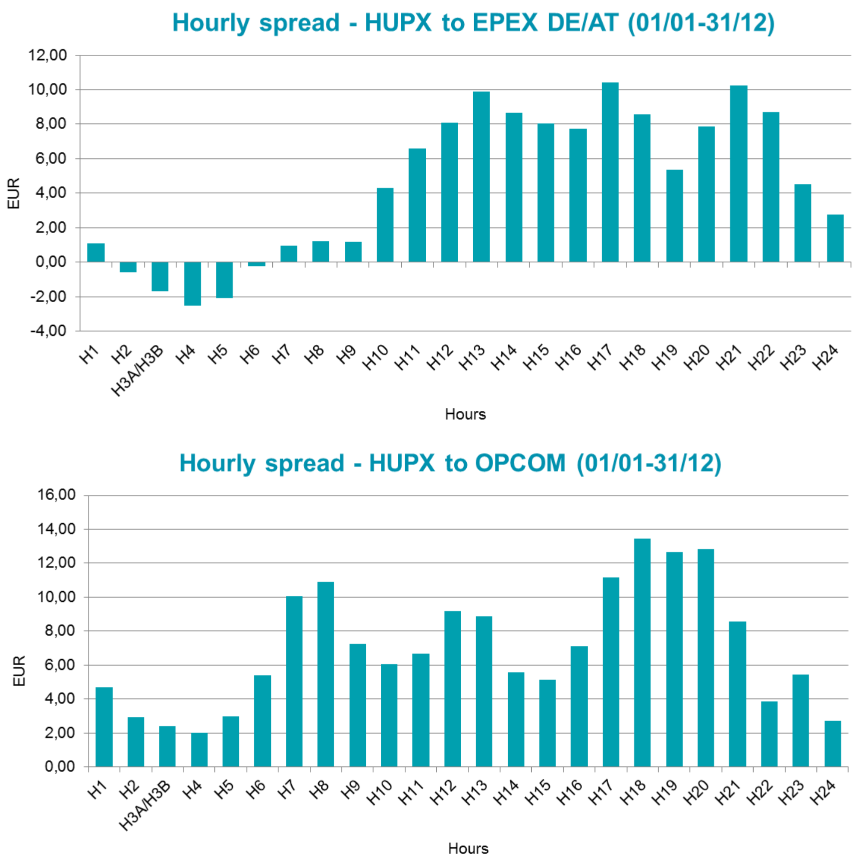
Source: IEA, *Energy Policies of IEA Countries, Hungary 2011 Review*, 2011, p. 97.

The HUPXDAM auction system is operating on the EPEX (European Power Exchange) trading system. The EPEX trading system is also used in Germany, Austria, Switzerland and France. The system enables submitting single hour and block hour bids. All bids must be submitted until gate closure at 11:00 a.m. Bids from the coupled area (HU, CZ and SK) are matched together and the spot price is determined on the basis of cross-border constraints. Bids can be placed with a price range from 3000.00 €/MWh to -3000.00 €/MWh. Auction results are published at 11:30 a.m. Day-ahead power trading in Hungary usually starts around 9:20 a.m. when day-ahead CAO (Central Allocation Office) capacity auction results are published. CAO holds auctions for AT-HU and HR-HU borders. Day-ahead power trading before HUPX gate closure is done at the OTC market. The OTC price on average reflects the HUPX spot price, but because many factors are influencing the HUPX spot price, the OTC price can also be misleading. The most liquid product sold on the OTC market is Base. This price difference between the spot and OTC market is an incentive for traders to develop models for price forecasting.

In 2013, the total traded volume on HUPX DAM was 9,074,022.9 MWh. In November the monthly traded volume reached a new record: 901,634.9 MWh. The highest daily traded volume was 41,008.6 MWh on 8 April delivery day, which is also the record daily traded volume since the launch of the HUPX DAM market. The average hourly trade volume was 1,035.8 MWh. The average daily traded volume on weekdays was 25,166.7 MWh and on weekends 24,091.6 MWh. By the end of the year, the number of HUPX DAM members reached 53 (HUPX, 2013).

The average daily traded volume difference between weekdays and weekends is negligible, but the price difference is not. This may imply that due to high start-up costs of thermo power plants, owners are willing to sell energy even below their short-run marginal costs in order to avoid start-up costs.

Figure 8. HUPX Prices vs. Neighbouring Exchanges – Hourly Spread



Source: HUPX, *HUPXDAM annual report 2013*, 2013, p.6.

Figure 8 plots yearly hourly average spot price spreads between HUPX and the neighbouring PXs for every single hour. We can see that HUPX spot hourly prices are on average higher than neighbouring PXs. On average, hours from 2–6 are cheaper compared to EPEX because there is a negative spread. EPEX DE/AT is a power exchange for spot trading in Austria and Germany. The Romanian OPCOM was on average cheaper in all hours, but protected with an export-import fee. In 2013, the average spot price for HUPX Base was 42.33 €/MWh. The highest traded price for a single hour was 250.01 €/MWh and the lowest traded price amounted to -30.09 €/MWh (negative price).

Negative prices are a price signal on the power wholesale market that occurs when a high inflexible power generation meets low demand. Inflexible power sources can't be shut down and restarted in a quick and cost-efficient manner. Renewables do count in, as they are dependent from external factors (wind, sun) (Epexspot, 2014b).

Figure 9. HUPX, Auction Aggregated Curve for the 21st Hour, 13 May 2014

Source: HUPX, *Market data*, 2014a.

In Figure 9 we can see the supply and demand curve formation in practice. For the 21st hour traded on 13 May 2014, the spot price was determined at 42.52 €/MWh and the traded volume was 1343.1 MWh. According to the merit order pricing (Figure 2), the proposed theory of microeconomics suggests that the first 1,000 MWh are supplied from renewable sources (negligible short-run marginal costs). The remaining volume to reach 1343.1 MWh is probably supplied by a nuclear or thermo power plant (fired on lignite or coal), i.e. from a power plant whose short-run marginal costs would be around 42 €/MWh. Nuclear power plants are not that appropriate for day-ahead trading since they have very limited production variability, which is why the majority of their production is hedged in advance. As a result, in the case of the 21st hour, the thermo power plant is probably the price setter. This is an explanation of the formation of the demand and supply curves based on the theory of microeconomics.

However, it is important to note that the market is not perfect and that the supply and demand curves are intertwined with trading strategies of pure traders seeking for a profit, and bidding strategies of generator owners. We can see that the visible part of the demand curve has a step function shape and is steeper compared to the supply curve, but not completely inelastic as proposed in Figure 3. The demand at the starting price is actually 3000.00 €/MWh around 300 MWh and is very steep (almost inelastic) before the visible part in Figure 9. The visible part of the demand curve in Figure 9 is not inelastic because at this section of the demand function, market participants are acting according to their trading strategies. This is why the shape in this part is a step function. Where the demand function has a step function shape, the demand could be considered as a speculative demand (i.e. power plants setting bids to replace their scheduled production according to their short-run marginal costs and pure traders seeking profit). Where the demand curve is almost inelastic, it could probably not have been forecasted in advance. Thus, it must be bought at any price at the day-ahead market to keep the power system in balance. The supply at the starting price –3000.00 €/MWh is around 360 MWh. This means that the suppliers are willing to pay the buyers to buy energy. This is probably due to the unpredictable nature of renewable production sources as the producers did not expect an increase in the production and were unable to sell energy in advance. This could also be their selling strategy because they hoped to negotiate a better price at day-ahead market compared to the prices that had been offered in the forward market. This is why they leave a part of their production to be sold at the day-ahead market. The intersection of the curves indicates that 200 MWh less supply with the price of 0.06 €/MWh would cause a new price settlement at around 48–50 €/MWh. This significant price swing can be mainly attributed to the low market depth[2]. As a result, relatively small changes on the supply and demand side cause a major price swing. Considering HUPX as a balancing point for the Balkan region, the bidding strategies of all market participants in the region are incorporated in these curves.

2 DATA

The previous chapter briefly discusses the factors influencing HUPX spot price determination. Considering all possible factors (fuel price, price of CO_2 coupons, unscheduled unavailability of power plants, congestion issues etc.), the effect on the spot

[2] If a market is deep, small changes in the demand and supply side do not cause a major price change.

price determination is a complex issue that is beyond the scope of this thesis. The purpose of this thesis revolves around price forecasting using the time series model and looking into the influence of the Balkan region on HUPX spot price determination. The data set will be based on publicly available data and supported with internal sources. The analysis will focus on the demand side and the supply side. Moreover, the data set will include the HUPX spot price, the daily average temperature, the realised wind production and the stream of the Drava River. Because the spot market sets a price for every single hour of the day, it is sensible to obtain all data for every single hour of the day. However, this was not possible for all variables because only daily averages were available. Hourly values for the river flow and temperatures were missing, hence all the values represent daily averages. In addition, there is a suggestion found in literature to rather work with daily averages when examining the effect of exogenous variables (e.g. wind and hydro production) on the spot price (Würzburg et al., 2013). Working with daily averages also prevents tedious work with hourly data sets on account of different time zones in the region.

The dependent variable in the econometric model is the HUPX spot price. The data set on the HUPX spot price has been available since July 2010 (the launch of HUPX). Since 12 September 2012, HUPX has been coupled with two power exchanges (OKTE and OTE). My data set will span the period from 21 September 2012 to 20 September 2013. The observations before 12 September 2012 will not be considered due to the new market structure (market coupling) launched on 12 September 2012. The installed hydro production capacities have been more or less fixed over time, but the observation of the statistics regarding the installed wind production capacities (EWEA, 2013) reveals huge variations in time. During the time span in question, Croatia added 122 MW and Romania 695 MW of wind production capacities. As the data set covers 1 year, all four seasons are included in the data set. This is important because every season has its own specifics.

The demand side in the model will be represented through the daily average temperature in the Balkans. The demand is hard to predict owing to different sectors influencing the consumption (Figure 7). The spot market serves mainly for balancing everything that could not have been predicted in advance. Temperature deviations from historical average values are hard to predict, which is why they affect the spot price determination. The majority of residential consumption comes from using heating and cooling appliances. Their usage is connected with temperature and GDP. On the one hand, higher summer temperatures result in using air conditioning appliances, while on the other hand, lower winter temperatures result in using heating appliances. Additionally, a higher GDP means higher energy efficiency because residents can afford more energy-efficient appliances and housing. On average, residential consumption connected to using other appliances remains constant throughout the year. Energy consumption in other sectors mainly depends on the quantity of orders, GDP, etc. and is not directly connected with temperature. Market participants are using temperature forecasts for day-ahead price forecasting. I was unable to obtain a data set with forecasted values. To this end, I used the available data set with measured temperatures. I am convinced that there is almost no difference between the measured and forecasted temperatures. The daily average temperature will be collected for capital cities in a specified region. Population density in city areas is higher and therefore temperatures measured in urban areas are a better proxy for consumption. Traders are using forecasted temperature values for price forecasting purposes, however, I was unable to obtain the forecasted values. To this end, I used the measured temperature values as a proxy for the forecasted values. The temperature data set is collected from historical weather web page (Wunderground, 2014)

The supply side will focus on the production from renewable sources due to negligible short-run marginal costs and their important effect on market dynamics. Hydro, wind and photovoltaic power plants are installed in the Balkan region. The production of photovoltaic power plants depends on available sun radiation. Cloud coverage, for example, reduces photovoltaic production significantly. I was able to obtain data on photovoltaic production only for Romania, but only from the second half of 2013 onwards. The installed capacity in the region as a whole is low (the majority in Romania, Serbia and Slovenia) and due to the unavailability of the data set on photovoltaic production for the region in the relevant time span, I decided to leave this variable out from the modelling part.

Wind power plants are installed in Croatia, Hungary and Romania (cumulative power of 3200 MW). I was unable to obtain the realised wind production data for Croatia, so only the data sets for Hungary and Romania are taken into account. The data set for the wind production can be obtained on TSO's web site, for example for Hungary (Mavir, 2014) and for Romania (Transelectrica, 2014). The only solution to replace the unavailable data set on the realised wind production in Croatia is to apply the average daily wind speed value for the region in Croatia in which the majority of wind farms in the country is installed. Wind speed is highly correlated with wind production, which is why it is the best available proxy. However, it is important to note that the installed power is changing over the year and this issue cannot be neglected. Therefore, only data for Romania and Hungary will be included in the forecasting models.

In Table 2 we can see that hydro production capacities represent a major share in the total installed production capacities in the Balkans (45% on average). This implies that the production in these countries crucially depends on the availability of hydro production. Due to a low HUPX market depth and huge hydro potential in region, hydro production should have a major impact on HUPX spot price determination. Unfortunately, I could not obtain the data set for participation or for river streams of all the major rivers in the region for the time period in question. The data set for participation could replace the data for river streams, but it would not capture the increase in the river stream during spring caused by snow melting. Figure 14 shows a vast river stream increase in spring that cannot be overlooked. I was only able to obtain the data sets for Slovenia. Although the data set for participation for Slovenia is available on the web site of the Slovenian Environment Agency, it will not be included because the data set on the river stream is more convenient. For the considered time span I obtained the data set for the Drava River that was measured in Slovenia (Dravograd). The Drava River runs through Austria, Slovenia and Croatia and then flows into the Danube River. The total installed hydro generation power on the Drava River is 1407 MW (Austria 574 MW, Slovenia 587 MW and Croatia 246 MW) (Reka Drava [River Drava] In *Wikipedia.*). The installed hydro production power on the Drava River represents roughly 8% of the installed power as seen in Table 2. Due to the abovementioned facts, the stream of the Drava River will be considered as a proxy for the Balkan region. The river stream in the available data set was measured in Dravograd. Dravograd can function as the starting point of this region. Therefore, the measured river stream in Dravograd becomes significant for the rest of the region with a time lag. If a data set could measure the Drava River stream in the middle of the region, i.e. in Croatia, we could capture the effect of tributary streams and the time lag bias would also be smaller. However, it would be best to obtain the data set of the river stream from as many possible measuring points and for as many rivers as possible. The data set for the Drava River stream was obtained for the years 2011&2012 from the web site of the National Meteorological Service of Slovenia (Arso, 2014). As for 2013, it was obtained from an internal source (HSE d.d., 2014a).

Table 2. Installed Hydro and Wind Production Capacities by Country

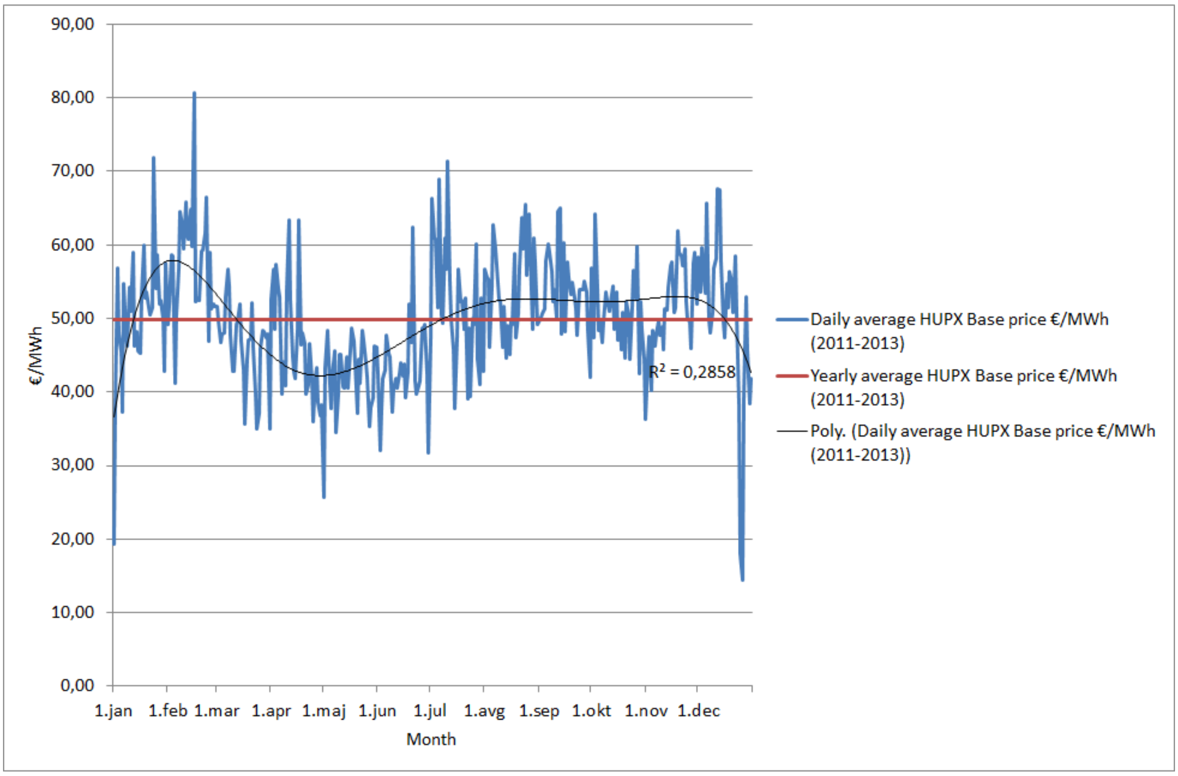
Country	**Hydro (MW)/% of total installed capacity**	**Wind (MW)/% of total installed capacity**
Bosnia and Herzegovina	2117/49,19	0/0
Croatia	1873/46,64	302/7,50
Hungary	53/0,58	330/3,60
Montenegro	658/75,81	0/0
Romania	6412/28,11	2599/11,40
Serbia	2225/24,82	0/0
Slovenia	1091/32,62	2/0,06

Source: Renewable facts (2012), *values for hydro production, &* EWEA (2013), *values for wind production.*

Figure 10 plots a daily average HUPX Base price over 3 years (2011–2013). A polynomial trend line of order 6 is added for a better perception of the trend. R^2 is a measure of how well the polynomial fits the actual data. The average HUPX Base price for the observed period was 49.88 €/MWh. Observing the spot price plot confirms previously described price drivers.

We can see the price peaking in winter and summer[3]. In winter, the highest price peaks are observed in February. In February, the spot price remained above the three-year average almost the entire the month. During that time, river streams are very low, while consumption increases due to heating. The lowest price can be spotted from 1 March until mid-June. Hydro production capacities in the Balkans represent roughly 45% of all production capacities. River streams during spring are at their peak due to melting snow, so the spot price is adjusted to this circumstance in line with expectations. Nevertheless, overall demand due to heating is falling because of rising temperatures.

Figure 10. HUPX Base Price 2011–2013



Source: HSE d.d. (2014b), *values for HUPX Base price.*

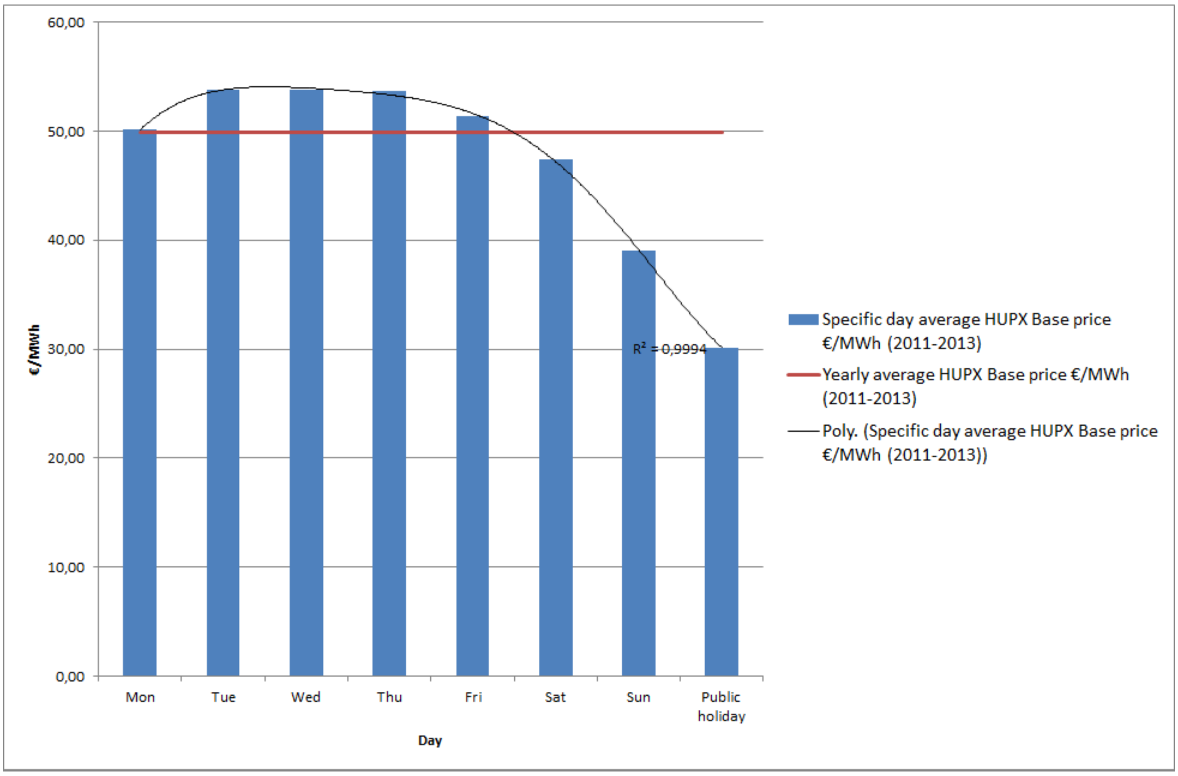
The price peak in summer is lower than in February; in addition, the longest period above the three-year average price is shorter compared to the one in February. At the end of summer, the price starts to fall; in October, prices even fall below the three-year average for a little while. In Figure 14 we can see that exactly at that time, the stream of the Drava River is peaking. The price drops significantly at the end of December. This fact attests to a significant consumption drop due to national holidays. This seasonality in price will be captured by using dummy[4] variables. The average HUPX Base spot price spanning from

[3] Electricity demand typically peaks in winter, but the summer peak is reaching closer to the winter peak level, mainly owing to the increasing use of air conditioning (IEA, 2011)
[4] Indicator variables are used to account for qualitative factors in econometric models. They are often called dummy, binary or dichotomous variables, because they take just two values, usually one or zero, to indicate

21 September 2012 to 20 September 2013 (the considered time span in further modelling) was 41.70 €/MWh; if we compare it to the three-year average as plotted in Figure 14, we can see that the base price is falling. In the period between 21 September 2012 and 20 September 2013, the minimum base price amounted to -6.71 €/MWh, while the maximum price reached 73.19 €/MWh.

Figure 11 reveals the specific day effect. Weekends and public holidays are clearly separated from weekdays. The lowest prices are recorded on public holidays. The average public holiday price is calculated for non-moveable public holidays in Hungary. A clear price fall on weekends and public holidays suggests using dummy variables to capture this specific day effect. This price drop is a consequence of lower demand because the majority of services and factories are closed on public holidays.

Figure 11. Average HUPX Base Price, by Day (2011–2013)



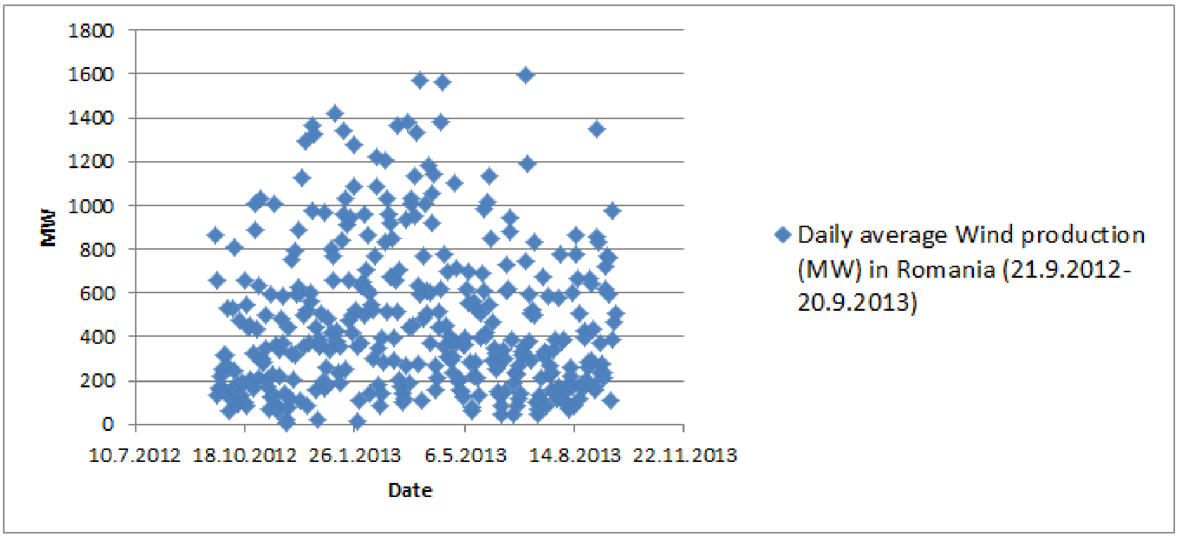
Source: HSE d.d. (2014b), *values for HUPX Base price.*

Daily average values for wind production were calculated for the observed time span (21 September 2012 to 20 September 2013). Some hours had negative values and were replaced by a zero. Wind production plots can be seen in Figures 12 and 13 for Romania and Hungary, respectively. The maximum value of daily wind production in Romania was 1600 MW, the minimum value amounted to 8 MW and the average to 479 MW. By the

the presence or absence of a characteristic or to indicate whether a condition is true or false (Hill, Griffiths, and Lim, 2011).

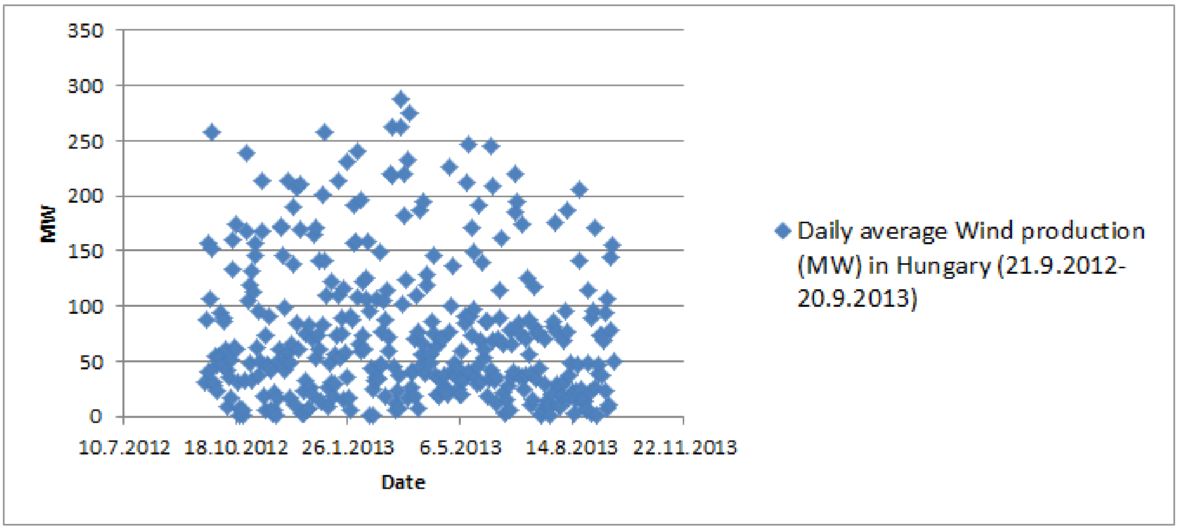
end of 2013, the installed wind production capacity in Romania was 2600 MW. Although the maximum measured value of production deviates for 1000 MW from the installed capacity, it is necessary to bear in mind that this maximum is comprised of 24 single-hour average productions. Even if some hours of the day are close to the installed maximum, these single hour maximums values are lost when taking the average of the day. The maximum value of the daily wind production in Hungary (Figure 13) was 287 MW, the minimum value 1 MW and the average value 76 MW. In Hungary the maximum value of 287 MW is much closer to its installed capacity of 330 MW than in Romania. Observing scatter plots does not reveal any patterns, meaning that wind production seems to be random. The plots of wind production are represented as scatter plots to enable easier visual inspection of densities. A higher density can be observed below 600 MW for Romania and below 100 MW for Hungary, which are approximately the average values of wind production. Since wind production is random, it should also affect the HUPX spot price on a random basis, without any seasonal particularities. When wind production is available, it should decrease the HUPX spot price. This effect will be examined more thoroughly in the modelling part.

Figure 12. Daily Average Wind Production in Romania



Source: Transelectrica (2014), *Values for wind production.*

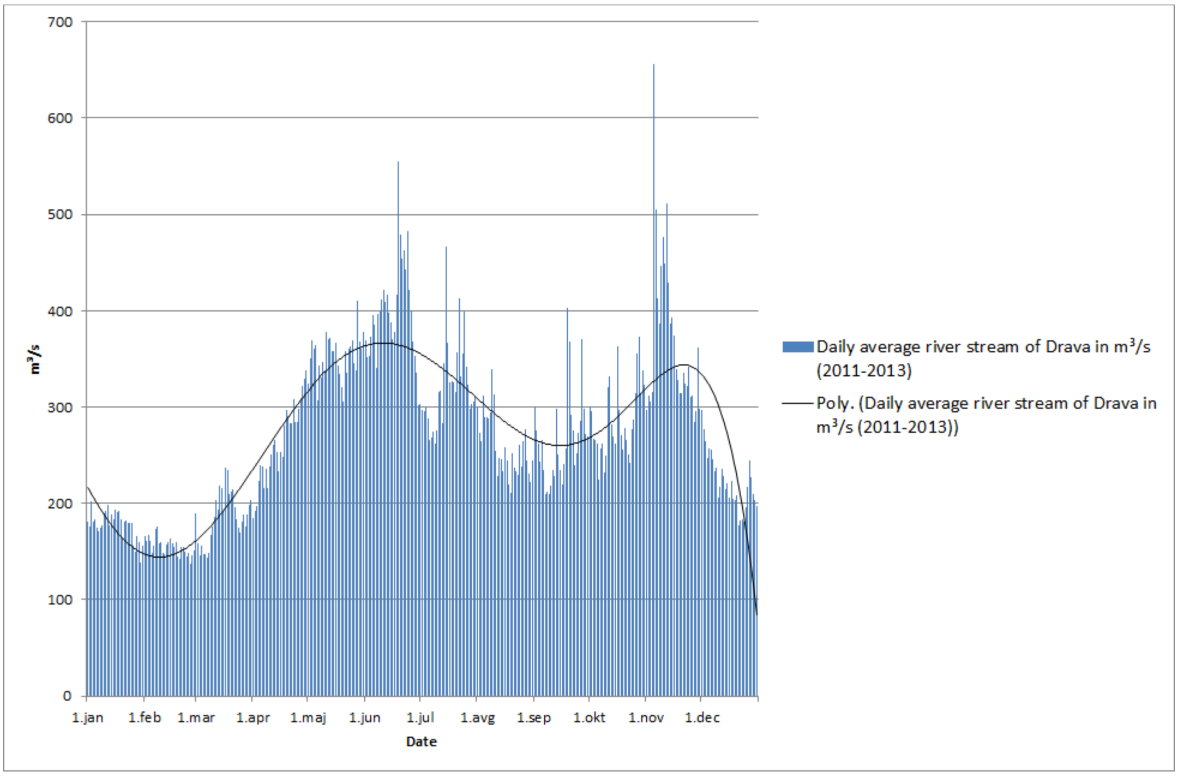
Figure 13. Daily Average Wind Production in Hungary



Source: Mavir (2014), *values for wind production.*

Figure 14 shows that the Drava River stream has 2 peaks. The first spans from April to July and the second from October to November. The first peak is connected to melting snow and ice in the Austrian mountains, and the second is a consequence of participation (Wikipedia, 2014b). The lowest river stream is recorded in January and February since participation in winter is mainly in the form of snow and accumulated in snow pack. The average river stream in the observed period was 272 m^3/s, the minimum 137 m^3/s and the maximum 656 m^3/s (floods in November 2011).

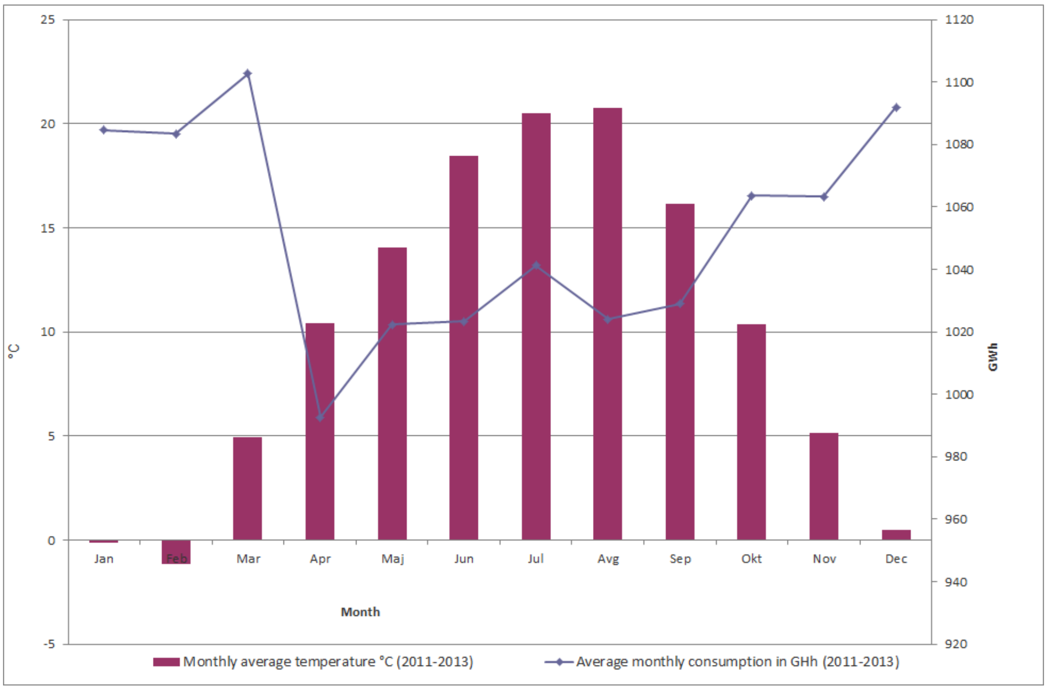
Figure 14. River Stream of the Drava River Measured in Dravograd



Source: Arso (2014), *values for river stream of Drava in years 2011&2012,* & HSE d.d. (2014a), *values for river stream of Drava in year 2013.*

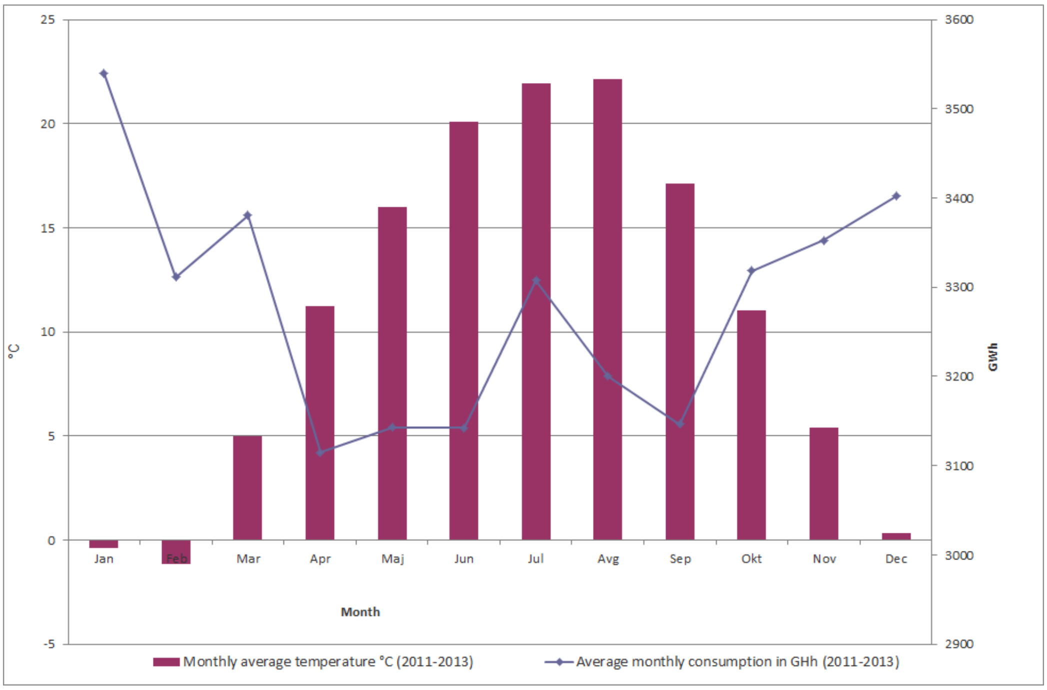
Figures 15–21 are plots of average monthly temperatures ($^{\circ}$C) in capital cities and average monthly consumption (GWh=Gigawatt-Hour) in a specific country. The data set on monthly consumption can be retrieved from the web site of the European Network of Transmission System Operators for Electricity (ENTSOE). Monthly average values for the period 2011–2013 are plotted in the Figures. These plots give us a better insight into the correlation between temperature and electricity consumption that was already discussed. We are mainly interested in the correlation in winter due to heating and in summer due to air conditioning. The Figures show that consumption profiles in the concerned countries have more or less the same shape. The difference in consumption levels is a consequence of different socio-economic factors (GDP level, electricity price to end consumers, population). All the observed countries show the highest consumption in winter on account of low temperatures (heating). In winter, temperature and consumption are negatively correlated. A more pronounced consumption peak in summer has not been observed in all countries. It can be spotted in Slovenia, Croatia, Hungary and in Montenegro. The abovementioned socio-economic differences in the region are clearly pronounced in the consumption in summer. Air conditioning can be considered as a luxurious good in countries with a low GDP. Hence there is no peak in the consumption in summer. Slovenia, Croatia and Hungary have the highest GDP in the region in question, so the peak in consumption is clearly visible. Montenegro is a popular holiday destination, which is why we can spot a consumption peak, although its GDP is relatively low compared to Slovenia, Croatia and Hungary. In Figure 17 we can see that Croatia has a consumption peak in summer that lasts throughout July and August. This peak is close to the level of the consumption in winter. Croatia is also a popular summer holiday destination and thus the consumption in summer increases. We can also notice that temperatures in the considered region are highly correlated. All these issues are considered in further modelling.

Figure 15. Temperature and Demand in Slovenia



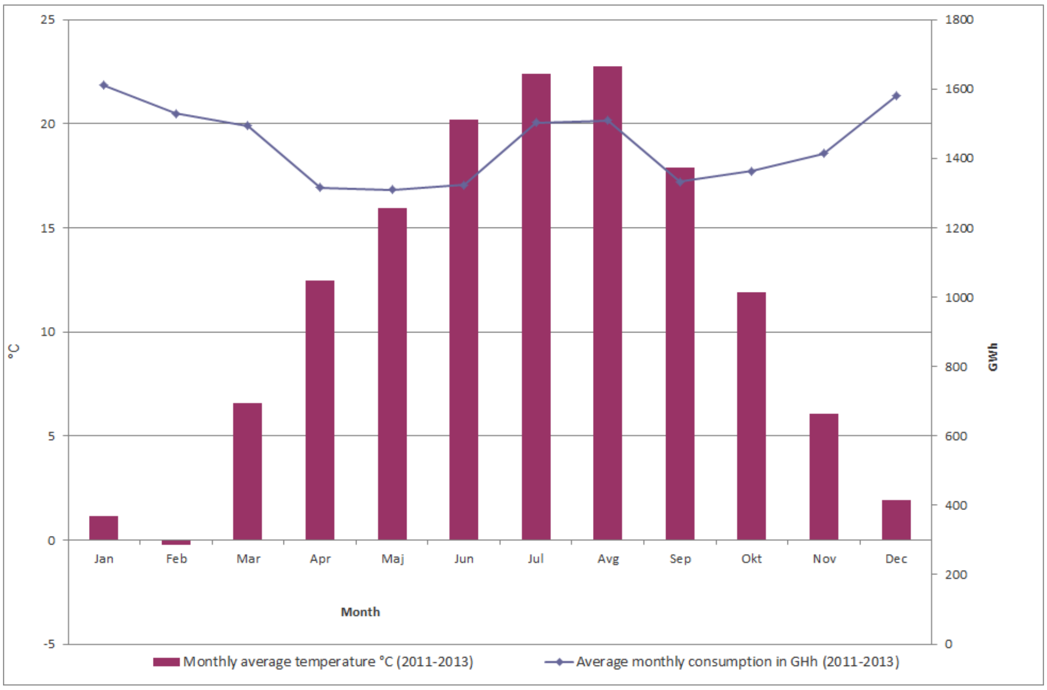
Source: Wunderground (2014), *values for daily temperature,* & ENTSOE (2014), *values for monthly consumption.*

Figure 16. Temperature and Demand in Hungary



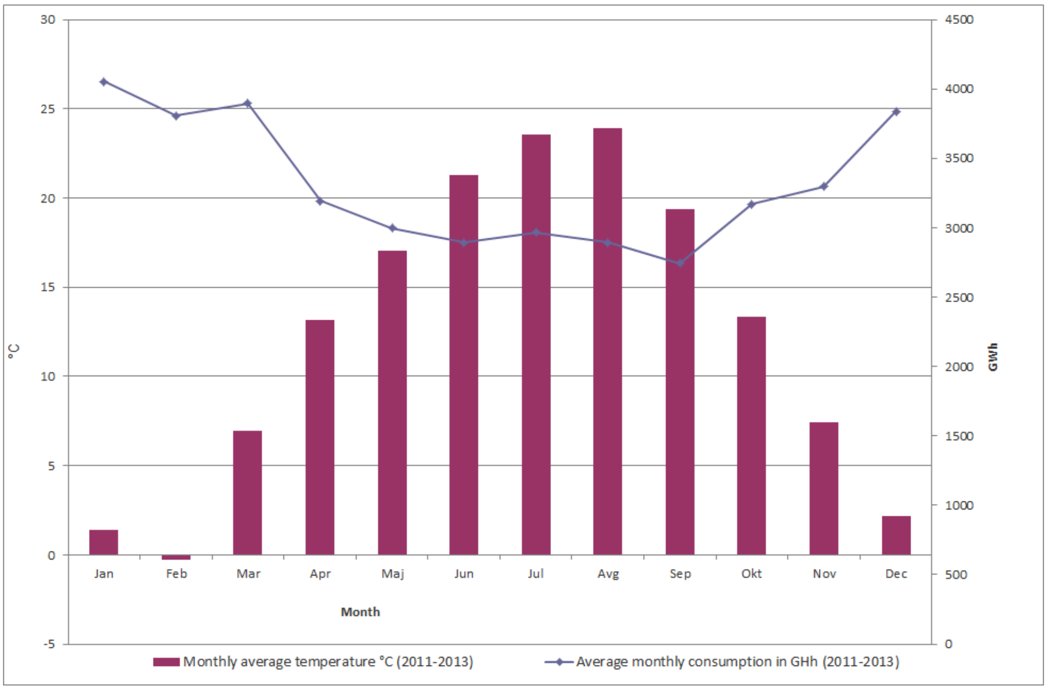
Source: Wunderground (2014), *values for daily temperature,* & ENTSOE (2014), *values for monthly consumption.*

Figure 17. Temperature and Demand in Croatia



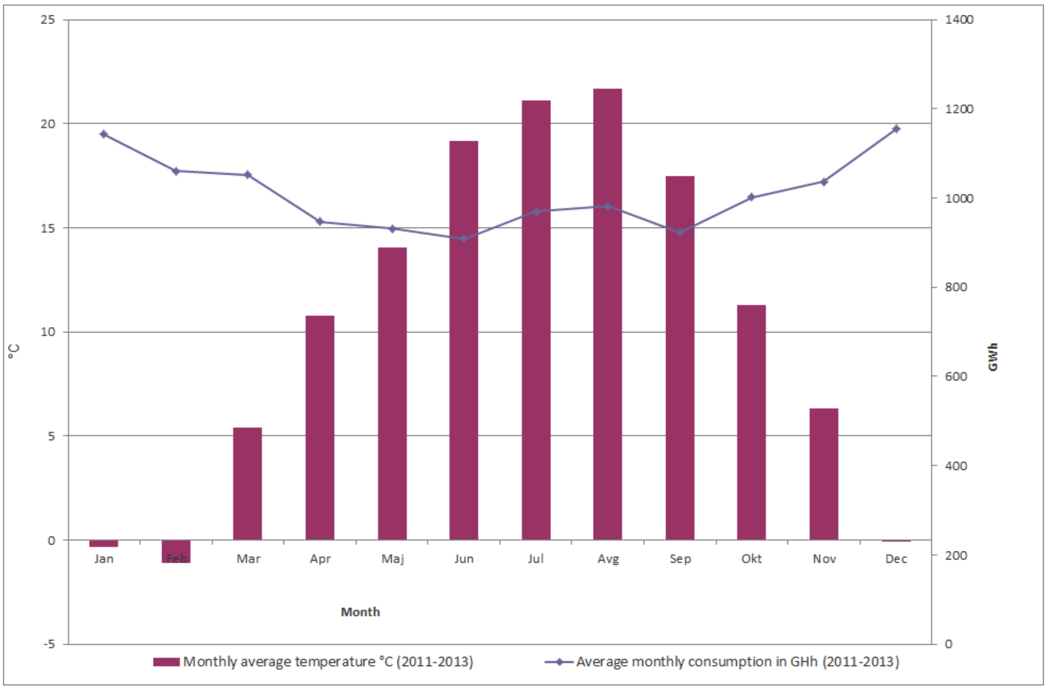
Source: Wunderground (2014), *values for daily temperature,* & ENTSOE (2014), *values for monthly consumption.*

Figure 18. Temperature and Demand in Serbia



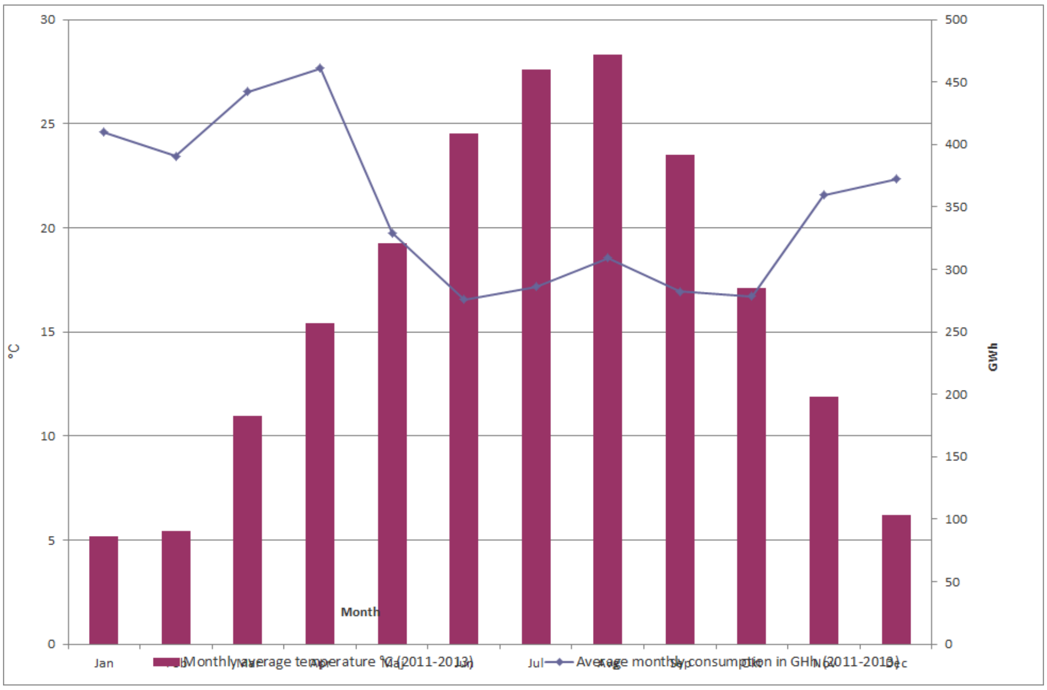
Source: Wunderground (2014), *values for daily temperature*, & ENTSOE (2014), *values for monthly consumption.*

Figure 19. Temperature and Demand in Bosnia and Herzegovina



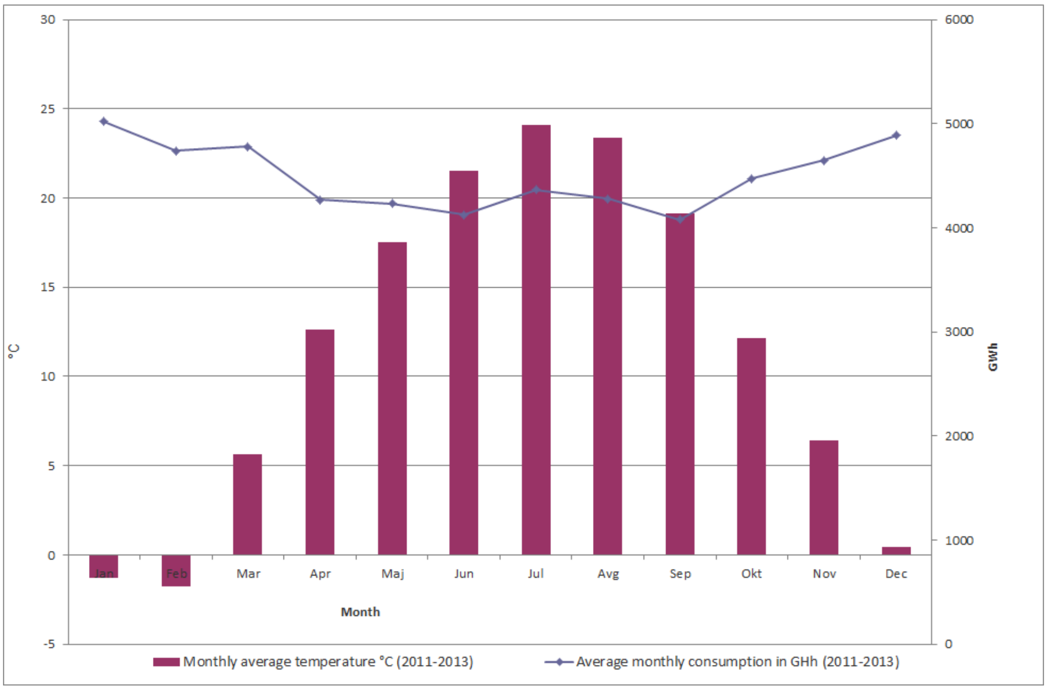
Source: Wunderground (2014), *values for daily temperature*, & ENTSOE (2014), *values for monthly consumption.*

Figure 20. Temperature and Demand in Montenegro



Source: Wunderground (2014), *values for daily temperature*, & ENTSOE (2014), *values for monthly consumption.*

Figure 21. Temperature and Demand in Romania



Source: Wunderground (2014), *values for daily temperature*, & ENTSOE (2014), *values for monthly consumption.*

3 METHODOLOGY

3.1 Explanation of approaches used in other papers

Literature proposes many different approaches to forecasting electricity prices. According to Aggarwal, Saini, and Kumar (2009), there are three main approaches available: game theory, simulation models and time series forecasting. Each of these main groups is further divided into several subsets.

The first group of models is based on the game theory. Analysis of the market equilibrium involves economics and game theory. In addition to the forecasted prices, this approach always come up with general equilibrium or market strategic behaviour (Li, Yu, Ren, Chiu, & Meng, 2013). It is very interesting to model the strategies (or gaming) of the market participants and identify the solution to those games. Since participants in oligopolistic electricity markets shift their bidding curves from their actual marginal costs in order to maximize their profits, these models involve the mathematical solution to these games, and price evolution can be considered as the outcome of a power transaction game. In this group of models, equilibrium models take the analysis of strategic market equilibrium as a key point. There are several equilibrium models available, such as Nash equilibrium, Cournot model, Bertrand model and supply function equilibrium model (Aggarwal et al., 2009).

The prediction of electricity prices based on a simulation model intends to solve a security constrained optimal power flow within the entire system range (Liu & Shi, 2013). These models form the second class of price-forecasting techniques, where an exact model of the system is built, and the solution is found using algorithms that consider the physical phenomenon that governs the process. Then, based on the model and the procedure, the simulation method establishes mathematical models and solves them for price forecasting. Price forecasting by simulation methods mimics the actual dispatch with system operating requirements and constraints. It intends to solve a security constrained optimal power flow (SCOPF) with the entire system range (Aggarwal et al., 2009). Although simulation-based price forecasting can provide a more detailed view of the price fluctuations, they require full insight into the system operation and hence are not practical for market participants (Zareipour, 2006). According to Skrivarhaug (in Duffner, 2012), fundamental models are used by utility companies as they have access to extensive datasets, e.g. Statkraft uses purely fundamental modelling in the spot market and forecasts the hourly dispatch for each of approximately 2500 modelled power plants in Europe. Simulation methods, also called fundamental models, have the capability to provide detailed insight into system prices, but these methods have two drawbacks. First, they require detailed system operation data and second, simulation methods are complicated to implement and their computational cost is very high (Aggarwal et al., 2009).

Time series econometrics is a rapidly evolving field. Generally, a time series is a sequence of values a specific variable has taken on over some period of time. The observations have a natural ordering in time (Krätzig & Lütkepohl, 2004). The time series forecasting methods use the past behaviour of electricity prices and some exogenous variables to forecast future electricity prices. In this group, two types of models are essential, and they are artificial intelligence techniques and conventional statistical models. Artificial intelligence techniques, such as artificial neural networks, are able to extract a nonlinear relationship governing inputs and outputs and then provide a prediction. Although artificial intelligence techniques can give an accurate prediction of electricity prices, one of its critical deficiencies is that the function forms built by them are implicit and the further analysis on the function forms such as sensitivity analysis is difficult (Liu & Shi, 2013). Many stochastic models are inspired by the financial literature and a desire to adapt some of the well-known and widely applied in practice approaches (Aggarwal et al., 2009). Univariate discrete type models widely used in practice are: autoregressive (AR), moving average (MA), autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA) and generalized autoregressive conditional heteroskedastic (GARCH) models.

The time series forecasting approach will be applied in the thesis. To this end, I will review some papers using time series approaches for price forecasting. None of the papers is considering the HUPX spot price forecast.

In the article "Extended ARMA Models for Estimating Price developments on Day-ahead Eectricity Markets" (Swider & Weber, 2007) the ARMAX approach is used for the analysis of the reserve market[5] price development in Germany. In the ARMAX model the authors use the German (EEX) spot market price as an exogenous variable. They propose further research focusing on: (i) incorporating other exogenous factors; (ii) applying the proposed models to out-of-sample forecasting; and (iii) evaluating their usefulness on other electricity markets.

The article "Forecasting Spot Electricity Prices with Time Series Models" (Misiorek & Weron, 2005) assesses the forecasting performance of the ARMAX model in the California power market. An exogenous variable considered in the model was residual load[6]. According to the authors, this model provides reasonably accurate price forecasts, at least for calm and moderately volatile periods. The authors found that the model for the

[5] As electricity cannot be stored in any major quantities, the amounts of electricity generated and consumed have to match exactly. Within a defined region, this system balancing is in the responsibility of a transmission system operator (TSO). The TSO must guarantee to have enough excess generation available for use at all times so that if, e.g. one generator fails, all loads may still be served without interruption (Swider & Weber, 2007).

[6] Residual load or residual production is determined by subtracting renweable infeed from demand (Wagner, 2014).

Californian power market performs the best with the inclusion of 3 dummy variables (Saturday, Sunday and Monday).

In the master thesis “Forecasting German Day-Ahead Electricity Prices Using Multivariate Time Series Models” (Duffner, 2012) the author uses the ARIMAX model with exogenous variables for wind and solar feed-in, the availability of generation capacities, fuel price, outside temperatures, etc. The data set used in the thesis involved two years of observations. The same data set was also used for ordinary least square regression (OLS) and for multivariate generalized autoregressive conditional heteroskedasticity (MGARCH) models. The author suggests that OLS has not been used exhaustively in former scientific works, while this work shows that OLS actually does not perform worse in forecasting and requires less computation time.

In the article “Mid-term Electricity Market Clearing Price Forecasting: A Hybrid LSSVM and ARMAX Approach” (Yan & Chowdhury, 2013), the authors use the ARMAX model combined with least squares support vector machine (LSSVM) for their forecasts. The article focuses on forecasting from 1 month to 6 months, which is useful for arranging maintenance schedules. The LSSVM module is first utilized to predict the electricity market clearing price. After that, the ARMAX module is utilized to predict the adjustments for each predicted electricity price values resulting from the LSSVM module. The model is calibrated for the PJM interconnected market in the USA.

The article “A Combined Modelling Approach for Wind Power Feed-in and Electricity Spot Prices” (Keles, Genoese, Möst, Ortlieb, & Fichtner, 2013) is dealing with wind power generation and its impacts on electricity price. Wind power generation and feed-in have a significant impact on electricity wholesale prices, especially for hours with high demand. The authors used the combined modelling approach for the simulation of wind power feed-in (WPF) series and electricity prices considering the impact of WPF on prices based on an autoregressive approach. The data set for Germany was used in the article.

The paper “Renewable Generation and Electricity Prices: Taking Stock and New Evidence for Germany and Austria” (Würzburg et al., 2013) exclusively deals with quantifying merit order effect on the spot price, thus identifying the costs and benefits of increased renewable capacity. They specified the multivariate model with dependent variable price change and explanatory variables: the demand for electricity, renewable production from solar and wind, the gas price, exports and imports of electricity and a set of dummy variables. They discovered that renewable production has a much higher impact on electricity prices when the electricity system is close to full capacity. In line with the theory of the merit order effect, the coefficient for the renewable production explanatory variable is negative and statistically significant. Their estimation revealed that each additional expected GWh of renewable sources (wind production and photovoltaic) on average decreases the day-ahead price for 1 €/MWh. The estimation was performed on German (EEX) prices.

The paper “The Impact of Wind Generation on the Electricity Spot-Market Price Level and Variance: The Texas Experience” (Woo, Horowitz, Moore, & Pacheco, 2011) is considering the pure effect of wind generation on 15-min (intraday) balancing energy market within each of the four ERCOT (electric grid operator in Texas) zonal markets. Marginal generation in the area is considered to be natural gas-fired. The dependent variable in the econometric model is the 15-min balancing-energy market price, wind generation, nuclear generation, system load, natural-gas price and a set of dummies. This model was used for each zone. All explanatory variables were statistically significant and interpreted according to economic reasoning. On average, the wind generation price decreases from 0.32 $/MWh to 1.53 $/MWh (depending on the zone).

In the article “Stochastic Factor Model for Electricity Spot Price – The Case of the Nordic Market” (Vehviläinen & Pyykkönen, 2004), the authors firstly model fundamentals affecting the spot price and then combine them into a market equilibrium model to form the spot price. Since half of the production is hydro-based, the hydrological situation strongly affects the available supply in the Nordic area. Hydro availability modelling is focused on participation, snow-pack and reservoir levels. Temperature in the model serves to capture the snow-pack melting effect. Electricity demand is modelled with a fixed component (due to industry), a temperature-dependent component (variable demand) and a noise term.

3.2 Definition of the ARMAX model

ARMAX is a time series model with an additionally included exogenous part. The model actually consists of 3 components: autoregressive part (AR), moving average part (MA) and exogenous part (X). The X stands for the additionally included exogenous variables into the univariate ARMA model. Because ARMA model is an univariate model, it cannot be used for exploring exogenous effects (e.g. temperature) on the dependent variable (spot price). Univariate models are models of only 1 variable, in the case of price forecasting, i.e. the price and its lagged values. The ARMAX model is considered as a multivariate model owing to the exogenous part. The ARMAX model is actually an extension of the widely used ARMA forecasting model and thus appropriate for exploring the exogenous effects on the dependent variable, i.e. the price.

AR stands for the autoregressive part of the model. In our case, the AR model fits electricity price on its own lagged values. In equation (2) we can see that an AR process y_t of order *p i.e. AR(p)* process may be written as

$$y_t = \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} + u_t, \tag{2}$$

where u_t is an unobservable zero mean white noise process with time invariant variance $E(u_t^2) = \sigma_u^2$ and α_i are fixed coefficients (Krätzig & Lütkepohl, 2004). In a more compact notation using a lag operator L ($L^j y_t = y_{t-j}$) it can be rewritten as in equation (3)

$$(1 - \alpha_1 L - \cdots - \alpha_p L^p) y_t = u_t \tag{3}$$

Or

$\alpha(L) = y_t$, where $\alpha(L) = 1 - \alpha_1 L - \cdots \alpha_p L^p$ is the characteristic polynomial.

The AR order specification criteria, i.e. how many lags p should be considered in the model, can be based upon different criteria (Akaike, Hannan, Quinn, and Schwarz). The AR model's coefficients are often estimated using ordinary least squares (OLS). If at a time t a shock occurs, i.e. $u_t \neq 0$, the effect of this shock never ends in the process (in the limit, the value goes towards zero if a process is stationary). Due to the nature of the process, a shock at time t will have an effect on all future y_{t+i} values ($i>0$).

MA stands for the moving average part of the model. If the process y_t can be represented as in equation (4)

$$y_t = u_t + m_1 u_{t-1} + \cdots m_q u_{t-q} \tag{4}$$

the process is called a moving average of order q ($MA(q)$) (Krätzig & Lütkepohl, 2004). The u_t is a zero mean white noise process with time invariant variance $E(u_t^2) = \sigma_u^2$. It can be rewritten in a compact form using the lag operator L as in equation (5)

$$y_t = (1 + m_1 L + \cdots m_q L^q) u_t \tag{5}$$

Or

$y_t = m(L) u_t$, where $m(L) = 1 + m_1 L + \cdots + m_q L^q$ is the characteristic polynomial.

The maximum likelihood (ML) estimation is commonly used to estimate MA models (Tsay, 2005). Identifying the MA order is done by plotting the autocorrelation function. The nature of MA models allows a shock, i.e. $u_t \neq 0$, to effect the dependent variable just for a q period in the future since when shock occurs, it is present in the model for q periods. Shocks are $u_t \sim WN$ ($WN = White\ noise$) $(0, \sigma_u^2)$, i.e. $E(u_t^2) = \sigma_u^2$.

The ARMA(p,q) model is a combination of AR-part and MA-part. The AR-part has the order p and the MA-part the order q, i.e. the observations are described using the p previous observed values $y_{t-z} \forall\ z \in \{1,2,\ldots,p\}$ and q previous disturbances $\varepsilon_{t-w} \forall\ w \in$

$\{1,2,\dots,q\}$ (Swider & Weber, 2007). The ARMA*(p,q)* model is formally written as in equation (6)

$$y_t = \hat{y}_t - \varepsilon_t = \sum_{z=1}^{p} \alpha_z\, y_{t-z} + \sum_{w=1}^{q} \beta_z\, \varepsilon_{t-w} + \varepsilon_t \tag{6}$$

where the first term on the right-hand side is AR-part and the second part is MA-part.

As already stated, ARMAX*(p,q,r)* is just an extension of the ARMA*(p,q)* model with exogenous variables. The letter *r* reveals how many additional exogenous variables (X) are included in the model ($x_{s,t} \forall\, s \in \{1,2,\dots,r\}$).

In equation (7) we can see model specification ARMAX*(p, q, r)*:

$$y_t = \hat{y}_t - \varepsilon_t = \sum_{z=1}^{p} \alpha_z\, y_{t-z} + \sum_{w=1}^{q} \beta_w\, \varepsilon_{t-w} + \sum_{s=1}^{r} \gamma_s\, x_{s,t} + \varepsilon_t \tag{7}$$

In our case, X is a vector of following exogenous variables:

1. daily average temperature (capital cities in the region);
2. daily average wind production (Hungary and Romania);
3. daily average Drava river stream.

The ARMAX model has an MA part that cannot be estimated using OLS, since the lagged error variables cannot be observed. The respective parameters α_z, β_w and γ_s can be estimated by maximizing the log-likelihood function (Swider & Weber, 2007). MLEs (maximum likelihood estimators) are based on a particular distribution (e.g. standard normal, Poisson, Logistic, Bernuolli) assumed to have generated the observed random variable—dependent variable or error term (Greene, 2000).

3.3 Maximum likelihood (ml) estimator

In likelihood theory, the roles of data (observed values of a random variable) and parameters are turned around. Given the observed values *y* of the variable, the likelihood function is as in equation (8):

$$I(\theta|y) = f(y|\theta). \tag{8}$$

In the likelihood context, the parameter θ is unknown, while the data are known. In line with the regression model, the parameters are to be estimated (Greene, 2000). The estimator of mean for a sample from normal mean is found following way.

As equation (9) shows the Gaussian of normal density function is of form:

$$f(y_i,\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{y_i-\mu}{\sigma}\right)^2}. \tag{9}$$

The likelihood function for normal density can be seen in equation (10):

$$l(\mu,\sigma|y) = \sum_{i=1}^{N} ln \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{y_i-\mu}{\sigma}\right)^2} = \sum_{i=1}^{N} ln \frac{1}{\sqrt{2\pi\sigma}} - \frac{1}{2}\sum_{i=1}^{N}\left(\frac{y_i-\mu}{\sigma}\right)^2. \tag{10}$$

The estimate of μ is found by taking the derivative of l with respect to μ, as in equation (11):

$$\frac{\partial l(\mu,\sigma|y)}{\partial \mu} = -\frac{1}{2}2\sum_{i=1}^{N}\left(\frac{y_i-\mu}{\sigma}\right)\left(-\frac{1}{\sigma}\right) = 0 \Rightarrow \hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} y_i \tag{11}$$

In the case of multivariate regression, the likelihood function for a density that contains *y* and *x* and parameters β must be derived. The density function for observing the sample *y* for given values of *x* and parameter set θ is $f(y|x,\beta)$ (Greene, 2000).

Assuming independence of variables as in equation (12):

$$f(y|x,\beta) = f_1(y_1|x_1,\beta) \times f_2(y_2|x_2,\beta) \times \dots f_N(y_N|x_N,\beta), \tag{12}$$

which gives the log-likelihood function as in equation (13):

$$l(\beta|y,x) = ln\prod_{i=1}^{N} f(y_i|x_i,\beta) = \sum_{i=1}^{N} l_i(\beta|y_i,x_i). \tag{13}$$

The maximum of this function is found by standard optimization techniques for multiple variables (Greene, 2000). The derivatives for each β_k and first order conditions as equation (14) shows are:

$$\frac{\partial l(\beta|y,x)}{\partial \beta_k} = \sum_{i=1}^{N} \frac{\partial l_i(\beta|y_i,x_i)}{\partial \beta_k} = 0, \forall\, k. \tag{14}$$

Divided by *N* as in equation (15), this condition is known as a sample statistic, with population counterpart:

$$E\left(\frac{\partial l(\beta|y,x)}{\partial \beta_k}\right) = E\left(\sum_{i=1}^{N} \frac{\partial l_i(\beta|y_i,x_i)}{\partial \beta_k} = 0, \forall\, k.\right) \tag{15}$$

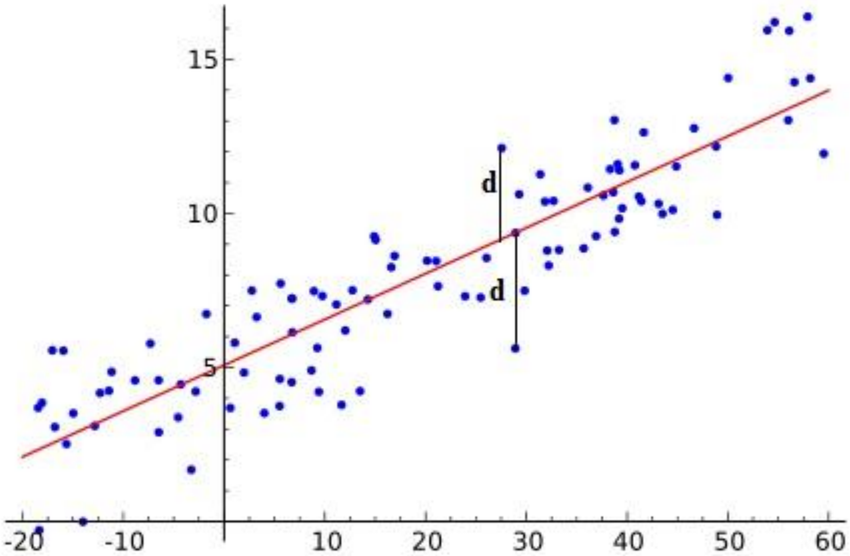
3.4 Ordinary least squares (ols) estimator

For estimating linear regression models, the least squares approach is considered as benchmark. The classical linear regression model is defined as a set of characteristics of the population that underlies an observed sample of data (Greene, 2000).

According to Greene (2000), these assumptions are:

1. linearity;
2. full rank;
3. exogeneity of the independent variables;
4. homoscedasticity and nonautocorrelation;
5. exogenously generated data;
6. normal distribution.

Figure 22. OLS Fitted Line



Source: Linear least squares. (n.d.) In *Wikipedia.*

Least squares principle asserts that to fit a line to the data values, we should make the sum of the squares of the vertical distances from each point to the line as small as possible (Hill et al., 2011). These distances (in Figure 22, denoted as d) are squared in order to prevent cancelling positive and negative distances. A line obtained using OLS fits the data points through the middle. Hence the tendency of such models is towards the mean of the observed data and spikes are unlikely to be predicted. This can be seen in Figure 22, where the fitted line goes through the middle of the observed data points.

3.5 Hypothesis testing and accuracy evaluation

The goal of the thesis is to check the influence of the Balkans (temperature, wind production, hydro production) on the market clearing price forecast. As already stated, these exogenous variables will be included as exogenous parts of the univariate time-series model and checked for statistical significance. If the additionally included exogenous variables turn out to be statistically significant and they have signs in line with the theory of economics, I will confirm the influence of the exogenous variables on spot price forecasts. Statistical hypothesis testing will be mainly based on t-tests, z-tests and chi-squared tests. It has become a standard practice to report a p-value of the tests. The p-value actually tells us to reject the null hypothesis if the p-value is less than or equal to the chosen level of significance α.

Accuracy of the estimated models for the out-of-sample forecast and the in-sample prediction can be measured with the mean absolute error (MAE), mean absolute percentage error (MAPE) and root mean square error (RMSE). These measures are widely used in papers dealing with time-series forecasting. The formulas are (Liu & Shi, 2013):

$$MAE = \frac{1}{T}\sum_{t=1}^{T}\left|P_{t,f} - P_t\right| \tag{16}$$

$$MAPE = \frac{100}{T}\sum_{t=1}^{T}\left|\frac{P_{t,f} - P_t}{P_t}\right| \tag{17}$$

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(P_{t,f} - P_t\right)^2}, \tag{18}$$

where $P_{t,f}$ is the predicted or forecasted price and P_t is the observed price.

We can see that the names of the error measurements are suggesting their characteristics by themselves. As equation (16) shows MAE is the average of the absolute errors. MAPE provides the same information, but it is expressed in percentages as in equation (17). If we compare RMSE and MAE, we should notice the quadratic term in equation (18), which amplifies measurement error with larger prediction deviations from observed prices. Since there is a rather insignificant difference between the proposed measurements, I will use MAE to compare pure time series models (ARIMA) with the extended ones for the exogenous part (ARIMAX).

4 EMPIRICAL RESULTS

All estimations, tests and data manipulations were conducted in Stata 12 and MS Excel. The estimations follow the Box-Jenkins procedure for time series models. According to Greene (2004), the Box-Jenkins approach consists of the following steps:

1. Satisfactorily transform the data so as to obtain a stationary series. This step will usually mean taking first differences, logs, or both to obtain a series whose autocorrelation function eventually display the characteristic exponential decay of a stationary series.
2. Estimate the parameters of the resulting ARMA (or ARIMA) model, generally by nonlinear least squares.
3. Generate the set of the residuals for the estimated model and verify that they satisfactorily resemble a white noise series. If not, respecify the model and return to step 2.
4. The model can now be used for forecasting purposes.

4.1 Data adjustment

The model includes the weighted average of temperatures (T_w) as equation (19) shows, since the correlation between measured temperatures in the region's capital cities is high (correlation matrix in the Appendix). The temperature in Romania is not included in the model, because Romania has an adequate, but protected PX. Temperature weights are calculated according to the share of yearly consumption in overall consumption in the considered region in the period 2011–2013.

$$T_w = T_{Lj}w_{Lj} + \cdots + T_{Bu}w_{Bu} \tag{19}$$

Table 3. Temperature Weights

City	**Weights**
Ljubljana	0,1
Belgrade	0,32
Sarajevo	0,1
Zageb	0,14
Podgorica	0,03
Budapest	0,31

According to Caro (2010), temperature transformation into heating and cooling indicators can be used. These indicators are tricky to use since a critical temperature for the cooling and heating mode needs to be chosen. On the other hand, the plots of the consumption and temperature in the entire region do not reveal a pronounced increased consumption due to air conditioning. The abovementioned plots reveal consumption peaks in the transition period from winter to spring (especially in March), which is probably a result of heating. In this transition period, the temperatures are varying around the critical values when heating is not necessary throughout the day. To this end, some households switch from primary heating sources to the most appropriate substitute, i.e. electricity, which causes this demand peak. Heating and cooling indicators will not be used since they require a deeper knowledge of heating and cooling habits in the region.

A consumption peak unusual for the region was spotted in Croatia and it lasted for almost the entire summer. For this reason and the peak in summer demand in Slovenia, Hungary and Montenegro, a dummy variable is included for the period from July to August. Having included this dummy variable, it will capture not only the effect of higher consumption in the abovementioned countries in summer, but also everything else that exogenously included variables fail to detect, but is statistically significant for this period.

The stream of the Drava River has a typical peak in late spring and early summer, otherwise it is stable. On the other hand, wind production is very volatile and requires transformation in order to eliminate volatility.

Many papers dealing with price forecasting used a logarithmic transformation in order to stabilize the variance. Taking logarithms is a monotonic transformation and thus has no effect on data. Since some variables also take negative values, a constant (100) is added to all variables when taking logs. Log-log regression form (i.e. $\ln(y) = \beta_1 + \beta_2 \ln(x) + \varepsilon$) is also straightforward when interpreting coefficients.

This equation is also known as the constant elasticity form as in the equation, the elasticity of y with respect to changes in x is $\frac{\partial \ln(y)}{\partial x} = \beta_2$, which does not vary with x (Greene, 2000). This means that if we change x by 1 percent, we would expect y to change by β_2 percent.

As already discussed in the data part, dummy variables are included to capture the special day effect. I will consider dummy variables for Saturdays, Sundays and public holidays. Holiday dummy variables are collected for all countries in the region except for BiH. Public holidays in BiH are omitted due to different jurisdictions and different religions in the country. In most of the cases, holidays in the considered countries are overlapping.

4.2 Autocorrelation

The correlation ρ between two random variables measures the degree of linear association between them (Hill et al., 2011). Autocorrelation is the correlation of a variable with its own past values. If a HUPX base price exhibits autocorrelation, this means that the price at time t+1 will be correlated with its past value in t. Variables and unobservable error term (ε_t) can be autocorrelated. The autocorrelation in the error term can arise from an autocorrelated omitted variable, or it can arise if a dependent variable y is autocorrelated and this autocorrelation is not adequately explained by the x*'s* and their lags that are included in the equation (Hill et al., 2011).

The correlation between two variables in our case y_t and y_{t-s} assuming that the time series is stationary, i.e.

1. $E(y_t) = \mu$
2. $var(y_t) = \sigma^2$, $cov(y_t, y_{t+s}) = cov(y_t, y_{t-s}) = \gamma_s$ for all $t \in T$ and integers s such that $t - s \in T$.

and the correlation is obtained as equation (20) shows

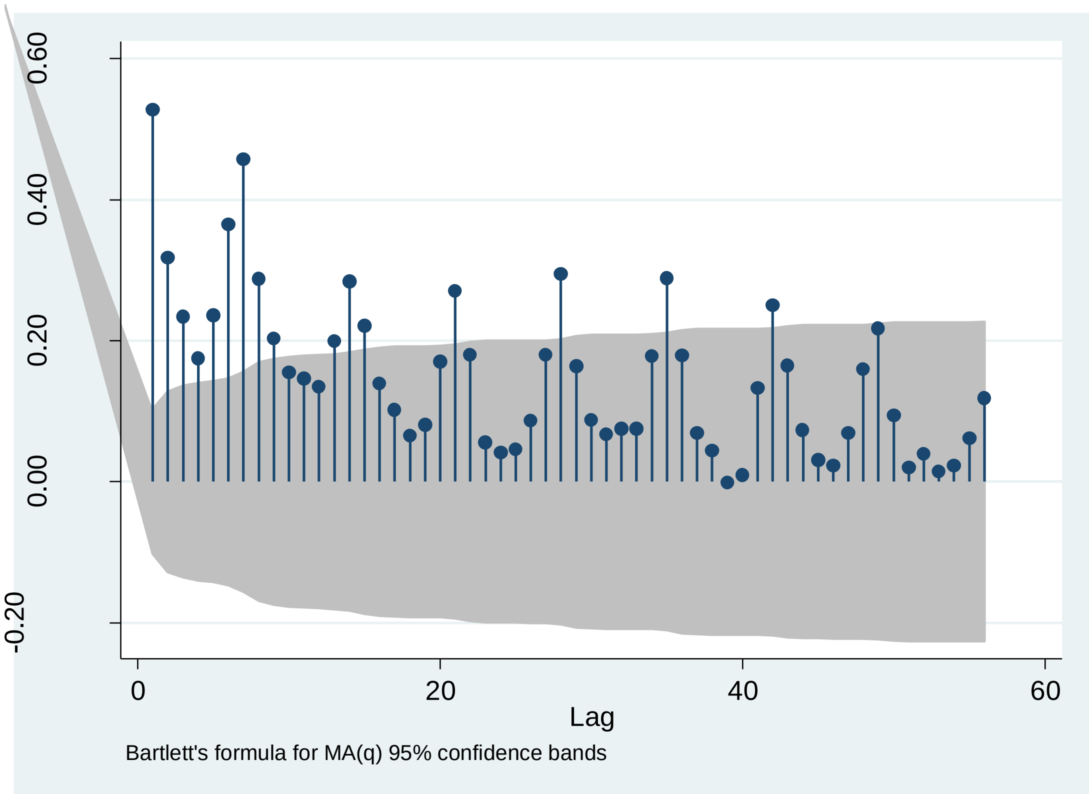
$$\rho_1 = \frac{cov(y_t, y_{t-s})}{\sqrt{var(y_t), var(y_{t-s})}} = \frac{cov(y_t, y_{t-s})}{var(y_t)}, \tag{20}$$

since $var(y_t) = var(y_{t-s})$.

In Figure 23 we can see the correlogram or a sample autocorrelation function for the HUPX spot base price. The correlogram shows the correlation between the observations that are *n* periods apart. The 95% confidence bands are represented with the shaded area. Autocorrelations that are lying outside the shaded area are significantly different from zero at a 5% significance level. In my case, I considered 56 periods (7 weeks). A pattern in the correlogram can be easily spotted. There is a strong correlation with a previous day and a day ahead. In addition, a statistically significant autocorrelation occurs in the 7-day period. This 7-day period is statistically significant even for 42 lags (6 weeks).

Statistically significant autocorrelation with the 7-day period suggests that the observed spot price on a specific day (e.g. Monday) in *t* will influence future Monday's price for almost 2 months (*t+42*). This is in line with the theory since each day has some specifics that are reflected in the spot price. This can be observed in Figure 11 where we can see that there is a difference in the price on different days of the week. The price starts dropping on Fridays and peaks on Mondays, whereas from Tuesdays to Thursdays it is stable.

Figure 23. Correlogram of the HUPX Spot Base Price



4.3 Stationarity

A stochastic process is called stationary if it has time-invariant first and second moments. These conditions are already stated in the autocorrelation part. The first condition means that all members of the stationary process have the same constant mean. Hence, a time series generated by a stationary stochastic process must fluctuate around a constant mean and does not have a trend. The second condition ensures that the variances are also time-invariant because for $s = 0$ the variance $\sigma_y^2 = E\left[\left(y_t - \mu_y\right)^2\right] = \gamma_0$ does not depend on t. Moreover, the covariances $E\left[\left(y_t - \mu_y\right)\left(y_{t-s} - \mu_y\right)\right] = \gamma_s$ do not depend on t, but just on the distance in time s of the two members of the process. The notation is also meant to imply that the means, variances and covariances are finite numbers, i.e. the first two moments and cross moments exist (Krätzig & Lütkepohl, 2004).

If the stochastic process is non-stationary (the characteristic polynomial has a unit root), the variance has a linear trend and therefore explodes as $t \to \infty$ *and* $s \to \infty$. Another issue arising from having unit roots is a spurious regression issue, since statistical inference is invalid when regressing non-stationary variables (Greene, 2000).

The autocorrelation function is a useful device for describing a time-series process in much the same way that moments are used to describe the distribution of a random variable. One of the characteristics of a stationary stochastic process is an autocorrelation function that either abruptly dies to zero at some finite lag or eventually tappers off to zero (Greene, 2000).

A formal test for testing a stationary is the Dickey-Fuller test. The null of the test is "a unit root is present in the series", hence rejecting the null means having a stationary time series. The AR(1) process $y_t = \rho y_{t-1} + u_t$ is stationary if $|\rho| < 1$. When $|\rho| = 1$ a process, a non-stationary random walk process is $y_t = y_{t-1} + u_t$. The Dickey-Fuller test is used to test the null hypothesis H_0: $|\rho| = 1$ against the alternative hypothesis H_0: $|\rho| < 1$.

After taking the first differences, I was able to reject the null hypothesis (non-stationary). More generally, the data generating process is said to be integrated of order $d(I(d))$ if first differences have to be applied d times to make the process stationary or asymptotically stationary (Krätzig & Lütkepohl, 2004). This means that the series is integrated of order 1 (i.e., $I(1)$).

The fact that the original time series is not stationary can be observed in Figure 10, where a 3-year daily average HUPX base price is plotted. It suggests that the first two moments are time dependent. The first moment (mean) is non-constant since seasonality in price is present (peak in winter, drop in spring etc.). The second moment, i.e. variance, is also non constant (higher deviations from mean in winter and summer). The correlogram in Figure 22 also points to this issue since the process needed 42 lags to die to zero. A plot of first differenced time series can be found in the Appendix 4, revealing that the time series becomes stationary after taking first differences. In the equation (7) originally proposed ARMAX(p, q, r) model is thus replaced with the autoregressive integrated moving average ARIMAX*(p,d,q,r)* model, where letter *d* stands for the order of integration. In our case is model written as ARIMAX (1,1,1,3). The model includes three exogenous quantitative variables. Dummy variables are actually qualitative variables, which is why they are not treated as a part of *r*.

4.4 Heteroscedasticity

Homoscedasticity is one of the assumptions of the classical linear regression model. It actually means that each disturbance ε_i has the same finite variance σ^2. Heteroscedasticity exists when the variances for all observations are not the same, i.e. $var(e_i) \neq \sigma^2$. According to Hill et al., (2011), there are two consequences for the OLS estimator when sample exhibits heteroscedasticity. Firstly, the OLS estimator is still a linear unbiased estimator, but it is no longer the best linear unbiased estimator (BLUE), because there is another estimator with a smaller variance. Secondly, the standard errors, usually computed for the least squares estimator, are incorrect, so the confidence intervals and hypothesis tests that use these standard errors may be misleading.

Heteroscedasticity may be detected by plotting residuals of the regression. If the errors are homoscedastic, there should be no patterns of any sort in the residuals and vice versa (Hill et al., 2011). There are also more formal tests to test for heteroscedasticity, such as the Lagrange multiplier test (Breusch-Pagan test), White test and Goldfeld-Quand test.

Most of the tests for heteroscedasticity are based on the following strategy: ordinary least squares is a consistent estimator of β even in the presence of heteroscedasticity. As such, the ordinary least squares residuals will mimic, albeit imperfectly because of sampling variability, the heteroscedasticity of the true disturbances. Therefore, tests designed to detect heteroscedasticity will, in most cases, be applied to the ordinary least squares residuals (Greene, 2000).

Hence the White test regress the explanatory variables x_i, squares of x_i and their cross-products on the least squares residuals $\widehat{e_i}$ obtained from the original regression $y = \beta_1 + \beta_i x_i + e_i$. If any of the explanatory variables is found to be statistically significant, heteroscedasticity is present in the sample. The White test is based on the chi-squared distribution.

If we are prepared to accept the least square estimator as a useful estimator, despite the fact that it is not the minimum variance estimator, there is a way of correcting the standard errors so that our interval estimates and hypothesis test are valid (Hill et al., 2011). The usage of robust standard errors is a solution to this issue. In large samples they are valid for heteroscedastic and also homoscedastic errors.

As Stata offers a specific option for robust maximum likelihood estimation, this function is always used in estimations. Hence the hypothesis testing is valid. For state-space models in general and ARMAX and ARIMA models in particular, the robust or quasi–maximum likelihood estimates (QMLEs) of variance are robust to symmetric non-normality in the disturbances, including, as a special case, heteroskedasticity (Stata).

4.5 Multicollinearity

The phenomenon when explenatory variables are highly correlated is referred to as multicolinearity. The case of the exact linear relationship among the regressors is a serious failure of the model, not the data. The more common case is one in which the variables are highly, but not perfectly, correlated. In this instance, the regression model retains all its assumed properties, although potentially severe statistical problems arise. The problem faced by applied researchers when regressors are highly, although not perfectly correlated include the following symptoms (Greene, 2000):

1. Small changes in the data produce wide swings in the parameter estimations.

2. Coefficients may have very high standard errors and low significance levels, even though they are jointly significant and the R^2 for the regression is quite high.
3. Coefficients may have the “wrong” sign or implausible magnitudes.

The severity of multicollinearity can be identified using the variance inflation factor (VIF) as equation (21) shows.

$$VIF = \frac{1}{tolerance} \tag{21}$$

$$tolerance = 1 - R_j^2 \tag{22}$$

R_j^2 in equation (22) is the coefficient of determination of a regression where *j-th* explanatory variable is regressed on all the other variables. VIF values above 10, i.e. tolerance below 0.1, are suggesting multicollinearity problems. The VIF is low for all the exogenous explanatory variables (mean VIF 1.33), so there is no problem with multicollinearity (detailed results can be found in the Appendix 7).

4.6 Estimation results

My first intension was to estimate two models, namely ARIMA (1,1,1) and ARIMAX (1,1,1,3), for full time span, i.e. 365 observations from 22 September 2012 to 21 September 2013, then check for statistical significance of exogenous variables and compare the prediction performance of the models. The ARIMAX model included weighted average temperature, the stream of the Drava River, wind production in Hungary and Romania, and dummies for: public holidays, Saturdays, Sundays, and the summer period. In the ARIMAX model for the full time span, none of the Balkan-based influences (hydro and wind production, temperature) turned out to be statistically significant ($p>0,05$). All the dummies were statistically significant ($p<0,05$), except for the dummy variable for holidays in Serbia and the dummy variable for the summer period, which were not (estimation output in the Appendix 7).

To this end, the sample was divided into four quarters (Q). These four quarters almost coincide with the four seasons (Q1 autumn, Q2 winter, Q3 spring and Q4 summer). The samples are almost balanced; Q1, Q2 and Q3 have 90 observations and Q4 has 95 observations. Furthermore, I summarized the data for each quarter in order to determine where exogenous variables reached their minimum and maximum values (summary statistics by quarter in the Appendix 8, 9, 10&11). I decided to divide the sample into four Qs due to the observed seasonality in the data section of the thesis. I expect exogenous variables in the individual quarters to be statistically significant, at least in the quarters where their values are most pronounced. For example, I expect for the wind production in Hungary and Romania to be significant in Q2 (winter) because both values are higher in this period compared to other quarters. The stream of the Drava River, i.e. hydro production, climbs to its maximum in Q1 and Q3. Predictably, the temperature achieved its maximum value in Q4 and minimum in Q2. Summary statistics indicate that there are more public holidays in Q2 and Q3, which is why I expect that dummy values for public holidays in these two quarters will be significant for more countries.

Table 4. Variables Definition

Variable	**definition**
ln_hupx	HUPX spot price, in €/MWh
ln_hydro	river stream of Drava, in m^3/s
ln_eol_hu	daily average wind production in Hungary, in MWh
ln_eol_ro	daily average wind production in Romania, in MWh
ln_w_t	daily weighted average temperature, in °C
d_sat	dummy variable for Saturdays
d_sun	dummy variable for Sundays
d_summer	dummy variable for summer
hol_hu	dummy variable for public holidays in Hungary
hol_cro	dummy variable for public holidays in Croatia
hol_slo	dummy variable for public holidays in Slovenia
hol_ro	dummy variable for public holidays in Romania
hol_sr	dummy variable for public holidays in Serbia
new_year	dummy variable for public holidays in the New Year period
ar	autoregressive term
ma	moving average term

For each quarter I estimated the ARIMA (1,1,1) model and the ARIMAX (1,1,1,3) model. The dummy variable for summer is omitted in the ARIMAX since each quarter is calibrated separately. I tried estimations with different numbers of lags for AR and MA parts, however, the samples were quite small so I was restricted in terms of trying different combinations of lags. Moreover, including an MA part with more than seven lags is usually statistically insignificant. In my opinion, it is reasonable to include an MA part with one lag because the tendency of the process is towards its mean. If we include an MA part with one lag, this means that the shock will be present in the model for one period. I find this reasonable because when a price spike occurs, the price in the next period usually turns back towards the mean of the process (this can be seen in Figure 24). So there is no reason to keep the shock in the model for more than one period. I tried to include different AR part lags (1, 7, 14) because Figure 23 was suggesting a high correlation with up to the 42^{th} lag. In ARIMA models, the AR part is statistically significant also for higher lags, however, in ARIMAX models, the number of lags must be lower.

To this end, I decided for estimating ARIMA (1,1,1) and ARIMAX (1,1,1,3) models to retain comparability. When estimating ARIMAX models, I first included all reasonable variables for an estimated quarter, then excluded all statistically insignificant variables and finally re-estimated the model. Although all specified variables are in line with the theory of economics, I decided to omit all insignificant variables also due to the over specification issue, because the estimated samples are relatively small (89 observations after taking first differences). On the other hand, the main goal of the thesis is not to find a perfect fit, but rather to explore the influence of the Balkans on price forecasting. The exclusion of statistically insignificant variables usually resulted in worse fit.

In the ARIMAX estimation for Q1, only the dummies for w**e**ekend and weighted average temperature are significant. All the signs are in line with the theory proposed at the beginning. The heating season starts in Q1 (autumn), so the negative sign for temperature is in line with the theory of economics. Higher temperatures are lowering the spot price, because heating-related consumption is lower. If the outside temperature changes by 1%, we are expecting the spot price to decrease by 0.42%. Figure 11 shows that on Sundays, the price drops more compared to Saturdays. This hypothesis is also confirmed by the signs and values of variables in Table 5. On Sundays, the spot price drops for 0.11% and on Saturdays for 0.04%. The P-value for AR part in the ARIMAX model is above the critical value ($p=0.05$). I ignored this fact because if a slightly higher significance level was chosen, i.e. $\alpha=10\%$, the p-value for AR part would remain below the critical value.

Table 5. Estimation Results for Q1

Variable	**ARIMAX**	**p-value**	**ARIMA**	**p-value2**
ln_w_t	-0,417971	0,018	/	/
d_sat	-0,038278	0,001	/	/
d_sun	-0,106876	0,000	/	/
AR(L1)	0,277828	0,088	0,35528	0,001
MA(L1)	-1,109368	0,000	-1	0,000

None of the public holidays turned out to be significant in Q1. In summary statistics for Q1 we can observe that this quarter contains less holidays than other quarters, but this should not be the reason for insignificance. I also checked whether public holidays coincide with the weekends, but this situation only occurred twice in Romania. Although the hydro production is high compared to other quarters, it is not significant in this quarter. In addition, wind production in Hungary and Romania turned out to be statistically insignificant in this quarter.

Figure 24 plots the actual and predicted price for Q1. It is easy to notice that the additionally included variables improved the fit; in the sample prediction, MAPE for

ARIMA is 15.85%, while it is 9.58% for ARIMAX. The inclusion of exogenous variables significantly improved the in-sample prediction. The values for error measurement are listed in Table 6.

Figure 24. Predicted and Actual Spot Prices for Q1 (In-sample Prediction)

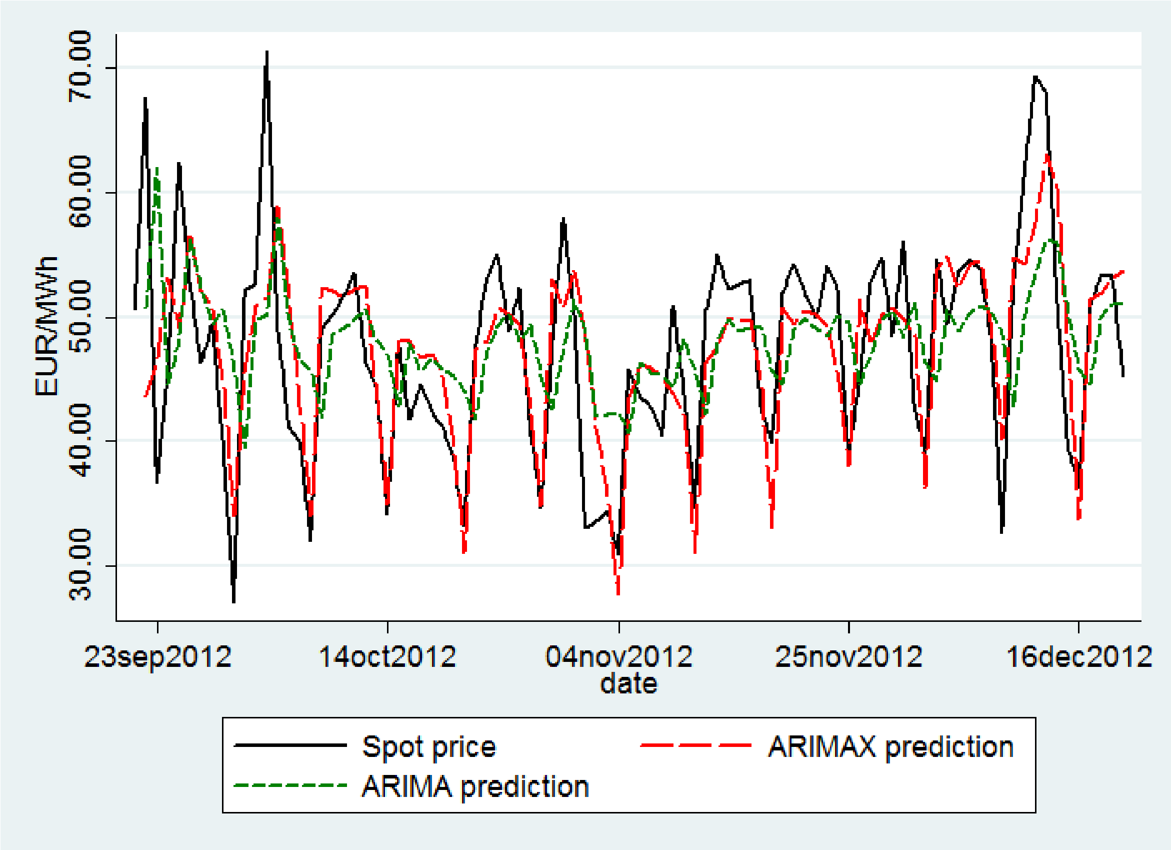
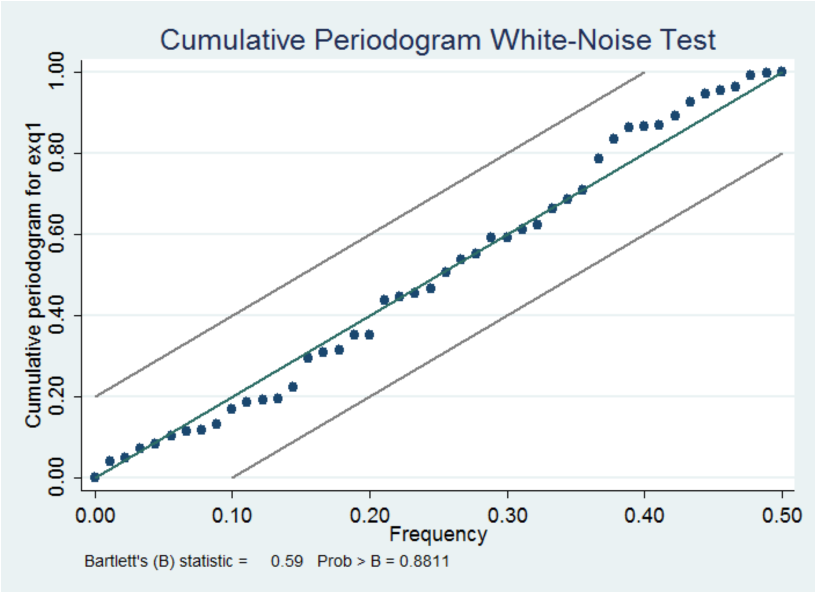


Table 6. MAE and MAPE for Q1.

Error measurement	**ARIMA**	**ARIMAX**
MAE(€/MWh)	6,96	4,50
MAPE (%)	15,85	9,58

According to the Box-Jenkins procedure, one has to check whether the generated set of the residuals for the estimated model resembles the white noise process. I assessed this using Bartlett's periodogram-based test for white noise. Bartlett’s test actually tests whether variances of two or more samples are equal, against the alternative hypothesis that at least two variances are not equal. In our case, the null hypothesis is that the residuals are white noise (i.e. variables are uncorrelated with constant mean and variance) and the alternative hypothesis that the residuals are not white noise. Figure 25 indicates that the process does not appear outside the critical bands. To this end, it can be concluded that the process is white noise and that the model can be used for forecasting purposes.

Figure 25. Test for White Noise Residuals for Q1



For Q2 I created an additional dummy variable *new_year* that takes the value of 1 on 25 and 26 December and on 1 January. Public holidays on these days are overlapping and a drastic drop in the spot price can be observed. To this end, I decided to create this variable to better capture this effect. From all the exogenously included variables, the following are statistically significant: hydro production (p=0,00), conditionally wind production in Romania (p=0,098), dummy variable for the New Year period (p=0,00), dummy variable for holidays in Croatia (p=0,014) and conditionally dummy variable for holidays in Slovenia (p=0,063). For price forecasting purposes I excluded hydro production and wind production because their signs do not comply with the proposed theory of economics; moreover, the AR term is statistically insignificant when they are present in the model. Both signs regarding production variables should be negative because these two production sources are the first on the merit order scale (negligible production costs) and hence their increased production should lower the spot price.

The fact is that both of the abovementioned power sources enter the Hungarian power grid through cross-border capacities. These capacities are assigned through explicit auction systems, and if the demand for transmission capacities for cross-border trading exceeds the available capacities, the auction price for capacities can be higher than the premium between two power exchanges. This could result in a higher HUPX spot price. In the case of increased hydro production in Slovenia, one could enjoy the price premium between BSP and HUPX since it is reasonable to expect that the BSP spot price would be lower than the HUPX spot price. If traders want to enjoy this premium, they have to secure cross-border capacities which can be overpaid, thus increasing the HUPX spot price. Hydro production in this period is at its minimum and the majority of participation is accumulated in snow pack. Higher temperatures melting this snow pack or rain can be over-anticipated by the market participants causing this situation. The case with wind production in Romania is more sensitive owing to its export fee (to my knowledge amounting to 4.3 €/MWh during the examined period), but reasoning could be the same, although wind production in Romania peaks in this period.

In the Appendix 16 is attached the correlation matrix between the spot price, hydro production and wind production in Q2. We can see that hydro production and the spot price are in fact negatively correlated, however, the spot price and wind production in Romania are positively correlated. This suggests that in the ARIMAX model the sign for hydro production should not be positive. Multicollinearity (low VIF values) is also not problematic, so the signs should not be "wrong". This issue should be thoroughly analysed, but it is beyond the scope of this research. As far as price prediction is concerned, I decided to exclude these two variables since I am not sure about the signs and the AR term is not statistically significant in their presence.

Dummy variables in this period for weekend days are not statistically significant. I checked whether the majority of public holidays overlapped with weekends, but this was not the case. The dummy created subsequently for the overlapping days is statistically insignificant. The temperature was also statistically insignificant, which could be attributed to the fact that this is a heating season and the majority of population is not using electricity for heating.

The ARIMAX model used for forecasting in Q2 is actually different from the ARIMA model just for the dummy variables for the New Year period and public holidays in Croatia. When hydro and wind production are excluded, public holidays in Slovenia become insignificant, however, the AR term becomes statistically significant.

MAPE for ARIMA is 34.65% and for ARIMAX 21.26%. ARIMAX again outperformed the ARIMA model in in-sample prediction. MAPE for ARIMAX with included wind and hydro production is 17.07%, meaning that a better fit is obtained even though the signs are not in line with the theory of economics.

It turns out that in Q3, the following variables were statistically significant: hydro production (p=0,001), temperature (p=0,016), dummy variables for Saturday (p=0,00), Sunday (p=0,00) and for public holidays in Hungary (p=0,00). All the signs are in line with the theory of economics, except for the sign for the temperature variable. If hydro production in Q3 increases for 1%, it reduces the HUPX spot price for 0.21%. According to the ARIMAX model, a higher temperature should increase the HUPX spot price. This is contrary to my expectations because by the end of this quarter, hydro production reaches its peak due to higher temperatures that result in a melting snow pack. Additionally, March, April and May can be considered as transitory months, during which electricity could sometimes function as a substitute for primary heating. Q3 ends on 17 June; during that period, temperatures are already high and an increase in consumption could be anticipated due to air conditioning. However, this should not justify a positive sign. The Appendix 24 contains a correlation matrix between the spot price, hydro production and temperature. Hydro production and temperature in Q3 are positively correlated and they are both negatively correlated with the spot price. The temperature variable was thus excluded from forecasting with the ARIMAX model.

The ARIMAX model again outperformed the ARIMA model. The calculated MAPE in-sample prediction is 17.22% for ARIMAX and 29.49% for ARIMA.

The Q4 sample has 95 observations, which is five more than in other quarters. In this quarter, the AR term in the ARIMA (1,1,1) model is statistically insignificant (p=0,165). As a result, I firstly omitted the AR part and used this new model. In this ARIMA (0,1,1) model, the MA part is statistically significant, but logically, the prediction power is very poor. Hence I applied the ARIMA (1,1,0) model and it turned out that the AR part is statistically significant. In sample prediction performance, this model is superior compared to the ARIMA (0,1,1) model. The ARIMA (1,1,0) model is very good in mimicking the dynamics of the process, whereas ARIMA (0,1,1) actually fits a positively sloped line. This means that the AR part is the main driver of the process and gives the shape of the process. If the MA part would be statistically significant in the ARIMA (1,1,1) model, it could be considered just as a corrector of the main driver of the process, i.e. the AR part.

Due to the above stated reasons, the ARIMA (1,1,0) model is a benchmark model. I tried with different lag combinations for the AR and MA part, but this model gives the best fit and contains a statistically significant part. The Appendix 32 contains an in-sample forecast done with ARIMA (0,1,1) and ARIMA (1,1,1) models for Q4.

ARIMA (1,1,0) is actually an AR (1) process where the first difference is taken in order to make the process stationary. Consequently, the MA part is also not included in the ARIMAX model and it is in the form of ARIMAX (1,1,0,3). From all exogenously included variables in the model, the following are statistically significant for this quarter: temperature (p=0,00), conditionally dummy variable for Saturdays (p=0,064), dummy variable for Sundays (p=0,007), and dummy variables for public holidays in Hungary (p=0,00), in Romania (p=0,00) and in Croatia (p=0,046). All the signs are in line with the theory of economics, i.e. dummy variables have negative signs, meaning that the spot price drops on public holidays and weekends due to lower consumption. As expected, temperature has a positive sign. If the temperature changes by 1%, the HUPX spot price increases by 1.09%. This number tells us that temperature in Q4 must be one of the decisive factors for spot price determination. One of the reasons must lie in air conditioning, which is used more during periods with higher temperatures. As we can see in Figure 14, the stream of the Drava River is declining in summer. Drought in summer decreases river streams and consequently hydro production. This situation is also connected with high temperatures and reflected in a higher spot price. The inclusion of exogenous variables in the ARIMA model improved in-sample prediction power since MAPE for the ARIMAX model is 19.42% and 26.04%for ARIMA.

As I was only able to obtain the data set for 2012 and 2013, the only option was to try the out-of-sample forecasting power of the calibrated model for the period from 21 September 2013 to 19 December 2013, in further analysis referred to as Q5. The spot price forecast with exogenous variables for Q5 is based on the ARIMAX model that is calibrated for Q1 in-sample prediction. Q1 and Q5 represent the same quarter of the year, hence it is normal to take the Q1 calibrated model for the out-of-sample forecast in Q5.

As already discussed, in the ARIMAX model calibrated for Q1, the following exogenous variables are included: temperature, dummy variables for Saturday and Sunday. When forecasting for the entire time span of Q5, the residuals turned out not to be white noise (see Appendix 35). This means that the residuals contained some information (innovations) that the model fails to explain. This means that this model should not be used for forecasting purposes. I visually checked the data and spotted an unusual spike on Sunday (27 October 2013). The HUPX spot price for Base was 81.7 €/MWh and that is incredibly high for a Sunday. In fact, this was the highest recorded price in all five Qs. I found on the Hungarian TSO's webpage (MAVIR) that on that particular day, there were two power plant outages in Hungary and such a high price must have been a consequence of the two events. Therefore I decided to forecast the spot price up to that event (36 predictions) and in this case, the residuals turned out to be white noise. Hence the ARIMAX calibrated for Q1 is suitable for forecasting purposes in Q5 up to this extreme event. The plot of the out-of-sample forecast for Q5 clearly shows how the two models, ARIMA and ARIMAX, predicted a price drop on 27 October (see Appendix 37).

MAPE for the out-of-sample forecast in Q5 is 17.79% for ARIMA and 13.68% for ARIMAX. Once again we can confirm that exogenously included variables result in a better fit. MAPE for the in-sample prediction in Q1 (36 predictions, the same time span as for out-of-sample forecast) is 15.85% for ARIMA and 10.59% for ARIMAX. As expected, the in-sample prediction gives a better fit, however, it can be concluded that the out-of-sample forecast for Q5 is still good compared to the in-sample prediction for Q1. The fit decreased by only 2% for the ARIMAX model and by 3% for the ARIMAX model.

4.7 Suggestions for model improvements

The above section provides proof that ARIMAX models have a much better accuracy in the in-sample prediction and also in the out-of-sample forecast compared to ARIMA models. As a result, it is important that exogenous variables included in the model are in line with the theory of economics for power markets. Also, the included variables must be statistically significant and they should be observed on as much as possible different places.

In our case, the stream of the Drava River is supposed reveal a clear picture of the overall hydro production in the Balkan region. As already mentioned, it is reasonable to assume that there is a correlation between river streams in the wider region because tributary streams flow into the main river (the Danube in the Balkans). However, we should not neglect the importance of local events (e.g. participations) since they increase the river stream in the main rivers with a time lag. In our case, the stream of the Drava River is measured in Slovenia. Consequently, we cannot assume that the overall river-run hydro production in the Balkan region is adjusted to the measured values in Slovenia. Therefore it would be good to have an extended data set of river streams for the considered region that is measured on as much as possible different places. This statement is justified in Table 2 showing that roughly 45% of the installed production capacities in the region is hydro production. Hence the variability in hydro production has a significant effect on the HUPX spot price.

Another issue that should be considered are "predicted price spikes with lag". For example, the prediction plot for Q4 (in the Appendix 33) shows that the tendency of the spot price is towards its mean. A spot price spike is immediately followed by a spot price drop and at that point, trouble starts for predicted spot prices. The problem is that right after the actual spot price spike, models predict a price spike at roughly the same level as the observed actual spike one period before. A simple solution to that problem would be to provide the model with a dummy variable that would capture this effect.

Some papers dealing with electricity price forecasting also include oil price, CO_2 coupons price, power plant outages etc. in the proposed models. Oil price is positively correlated with the price of fossil fuels, hence it is a good indicator for a dispatch of fossil fuel-fired power plants. The price of CO_2 coupons is relevant only for the producers within the European Union (EU). Due to the fact that the majority of the Balkan countries are not EU members, the inspection of CO_2 coupons price effect on the HUPX spot price would be interesting. Unplanned power plant outages of larger units are directly correlated with price spikes. A clear case of unplanned power plant outage is a price spike on Q5 Sunday 27 October. If it would be possible to obtain outage information prior the PX gate closure, one would be able to make a better price prediction. The inclusion of power plant outages could be done with a class of indicator variables for different levels of outages (i.e. 0–50 MW, 50–100MW etc.). To some extent, this would probably resolve the problem of "predicted price spikes with lag" discussed above.

It is logical that the in-sample prediction gives a better fit than the out-of-sample forecast because the model is calibrated on the basis of in-sample data. If we would like to use these models for creating bidding strategies, we would be purely interested in out-of-sample forecasts. In the case of out-of-sample forecast for Q5, the residuals are not white noise for the entire quarter (only for 36 predictions). A question arises about an optimal calibration period that yields the best fit for the out-of -sample forecast and how often should the model be recalibrated (every day, every week etc.) Econometrics electricity price forecasting is a very complex and demanding task as it is hard to incorporate all the relevant information proposed by the theory of economics in the models. If an extensive and detailed data set is available, a simulation approach is more suitable; however, it is time consuming and expensive.

CONCLUSION

The thesis is comprised of two parts. The first part contains an analysis of the main HUPX spot price drivers for which the data set is available and that can be reasonably included in forecasting models. The second part involves HUPX spot price forecasting using time-series models. The thesis is written in line with a systematic approach towards time-series modelling and deals with all the issues connected with time-series modelling. Price forecasting is done with univariate time-series models and multivariate time-series models. The univariate ARIMA model with the additionally included price drivers, analysed in the first part, is extended into the multivariate ARIMAX model, where X stands for the exogenously included variables. The advantage of being able to include exogenous variables into the ARIMAX model enables to statistically verify the influence of the Balkan region on the HUPX spot price forecast.

The daily average temperature in the Balkan region is statistically significant in autumn, spring and summer. In autumn, a higher temperature in the Balkans decreases the HUPX spot price, i.e. the model attaches a negative sign to the temperature. In spring, temperature is statistically significant, but a positive sign, i.e. a price increase when the temperature is higher, is not in line with the proposed theory in the ARIMAX model, because a higher temperature should decrease the HUPX spot price. The inspection of the correlation between the HUPX spot price and the daily average temperature reveals that the correlation in spring is indeed negative. In summer, a positive sign in the ARIMAX model, i.e. a price increase when temperature is higher, can be justified due to the increased electricity consumption caused by air conditioning.

The daily average wind production in Hungary has not been statistically significant for any of the examined quarters. As for the wind production in Romania, it turns out to be statistically significant in ARIMAX only in winter, which is when it peaks. However, a positive sign, i.e. price increase when the wind production is higher, is not in line with the theory, because it envisages a price drop on the days with a higher wind production. The examination of the correlation matrix for the considered period reveals that the HUPX spot price and daily average wind production in Romania are indeed positively correlated. Hence this question remains open and available for further analysis.

Hydro production, i.e. from the Drava River stream, is statistically significant in winter and spring for the ARIMAX model. By the end of spring, the stream of the Drava River reaches its peak, so the proposed negative sign, i.e. price drop when the river stream is higher according to the ARIMAX model, is in line with the theory. In winter, when the stream of the Drava River is at its minimum, the proposed positive sign, i.e. a price increase when the river stream is higher according to the ARIMAX model, is not in line with the theory of economics. On the days with a higher river stream, the HUPX spot price should drop and not increase. The examination of the correlation matrix reveals that the HUPX spot price and river stream of the Drava River are negatively correlated in winter.

The included dummy variables for weekends and for public holidays in the ARIMAX proved to be statistically significant in all quarters, except for the dummies for weekend days in winter, which were statistically insignificant. According to the model, the attached negative sign, i.e. a price drop, is in line with the proposed theory because lower HUPX spot prices could be observed on those days.

Statistical significance of the exogenously included variables in the ARIMAX model in the individual quarters , by the model attached signs that are in line with the proposed theory, confirm that the HUPX can be considered as a balancing point of the Balkans. Power systems must be in balance at all times. Hence the surpluses and shortages in the system must be balanced somewhere. Thus, HUPX can be considered as a one of the major balancing points in the Balkans.

The accuracy of in-sample price prediction and out-of-sample price forecasting improves significantly with multivariate models compared to univariate models, meaning that exogenously included variables improve accuracy. The calculated MAPE for the ARIMAX model is in all quarters lower compared to MAPE for ARIMA models. The lowest MAPE of 9.58% was calculated in Q1 for ARIMAX in-sample prediction, while MAPE in Q1 for the in-sample prediction with ARIMA model was 15.85%. The calculated MAPE for the out-of-sample forecast in Q5 based on the calibrated ARIMAX model for Q1 was 13.68%. The accuracy of these models for out-of-sample forecast is not bad compared to in-sample prediction.

None of the relevant papers regarding electricity price forecasting deal with the Hungarian Power Exchange (HUPX). Due to that fact, HUPX can be considered as one of the least explored power exchanges and therefore this thesis contributes to a better understanding of the major price drivers and provides a convenient time-series forecasting approach.

REFERENCE LIST

1. Aggarwal, S. K., Saini, L. M., & Kumar, A. (2009). Electricity price forecasting in deregulated markets: A review and evaluation. *International Journal of Electrical Power & Energy Systems, 31*(1), 13–22. Retrieved May 7, 2014, from http://www.sciencedirect.com/science/article/pii/S0142061508000884
2. Arso. (2014). *Arhiv hidroloških podatkov* (River stream of Drava in year 2012). Retrieved May 9, 2014, from http://vode.arso.gov.si/hidarhiv/pov_arhiv_tab.php
3. Bojnec, Š., & Papler, D. (2012). Renewable sources of energy: hydro-electricity in Slovenia. *Hrčak Portal znanstvenih časopisa Republikie Hrvatske.* Retrieved May 8, 2014, from http://hrcak.srce.hr/index.php?show=clanak&id_clanak_jezik=137685
4. BSP-SouthPool. (2013). *Market coupling.* Retrieved May 7, 2014, from http://www.bsp-southpool.com/market-coupling-siit.html
5. Cabrera, B. L., & Schulz, F. (2014). Probabilistic forecasts of electricity spot prices using residual load. Retrieved May 7, 2014, from https://lvb.wiwi.hu-berlin.de/research/veranstaltungen/hejnice-2014/talks/Schulz.pdf
6. Caro, G. (2010). Two empirical methods for forecasting spot prices and constructing price forward curves in the Swiss power market. Retrieved May 8, 2014, from http://www.cepe.ethz.ch/education/CompletedThesis/2010_05_Caro_Master_Thesis.pdf
7. Conejo, A. J., Contreras, J., Espínola, R., & Plazas, M. A. (2005). Forecasting electricity prices for a day-ahead pool-based electric energy market. *International Journal of Forecasting, 21*(3), 435–462. Retrieved May 5, 2014, from http://www.sciencedirect.com/science/article/pii/S0169207004001311
8. Duffner, S. (2012). Forecasting German day-ahead electricity prices using multivariate time series models. Retrieved May 5, 2014, from http://brage.bibsys.no/xmlui/bitstream/handle/11250/169717/Duffner%202012.pdf?sequence=1
9. Epexspot. (2014a). *OKTE, HUPX, OPCOM and EPEX SPOT to implement the PCR solution in the 4M MC project.* Retrieved May 5, 2014, from https://www.epexspot.com/en/press-media/press
10. Epexspot. (2014b). *Negative prices.* Retrieved August 8, 2014, from http://www.epexspot.com/en/company-info/basics_of_the_power_market/negative_prices
11. European Network of Transmission System Operators for Electricity (ENTSOE). (2014). *Load and consumption data.* Retrieved June 1, 2014, from https://www.entsoe.eu/data/data-portal/consumption/Pages/default.aspx
12. European Wind Energy Association (EWEA). (2010). *Wind energy and electricity prices.* Retrieved May 6, 2014, from http://www.ewea.org/fileadmin/ewea_documents/documents/publications/reports/MeritOrder.pdf
13. European Wind Energy Association (EWEA). (2013). *EWEA annual statistics 2013.* Retrieved May 20, 2014, from

http://www.ewea.org/fileadmin/files/library/publications/statistics/EWEA_Annual_Statistics_2013.pdf
14. Fezzi, C. (2007). *Econometric models for the analysis of electricity markets.* Retrieved June 5, 2014, from http://amsdottorato.unibo.it/433/1/tesi_dottorato_carlofezzi.pdf
15. García-Martos, C., Rodríguez, J., & Sánchez, M. J. (2012). Forecasting electricity prices by extracting dynamic common factors: application to the Iberian Market. *IET Generation, Transmission & Distribution, 6*(1), 11–20. Retrieved May 5, 2014, from http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=6128765&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D6128765
16. Genoese, M., Ragwitz, M., & Sensfuß, F. (2007). The Merit-order effect: A detailed analysis of the price effect of renewable electricity generation on spot market prices in Germany. *Fraunhofer ISI.* Retrieved May 9, 2014, from www.isi.fraunhofer.de/publ/downloads/isi07a18/merit-order-effect.pdf
17. Greene, W. H. (2000). *Econometric Analysis*. New Jersey: Pearson Education, Inc.
18. Gubina, A. (2010). Električna energija kot tržno blago [Electricity as a commodity]. Retrieved May 9, 2014, from http://lest.fe.uni-lj.si/sites/default/files/Gubina.Trg_.1.del_.Vodenje.10-11.lest_.pdf
19. Harris, C. (2006). *Electricity Markets*. West Sussex: John Wiley & Sons Inc.
20. Hill, R. C., Griffiths, W. E., & Lim, G. C. (2011). *Principles of econometrics.* Hoboken: John Willey & Sons, Inc.
21. Holding Slovenske Elektrarne d.d. (HSE). (2014a). *River stream of Drava in year 2013* (internal database). Ljubljana: HSE d.d.
22. Holding Slovenske Elektrarne d.d. (HSE). (2014b). *HUPX spot price in years 2012&2013* (internal database). Ljubljana: HSE d.d.
23. Huisman, R., Huurman, C., & Mahieu, R. (2007). Hourly electricity prices in day-ahead markets. *Energy Economics, 28*(2), 240–248. Retrieved May 12, 2014, from http://www.sciencedirect.com/science/article/pii/S0140988306000995
24. Hungarian Power Exchange (HUPX). (2013). *HUPXDAM annual report 2013.* Retrieved May 20, 2014, from http://www.hupx.hu/media/HUPX_DAM_ANNUAL_REPORT_2013_PUBLIC_404.pdf
25. Hungarian Power Exchange (HUPX). (2014a). *Market data*. Retrieved May 7, 2014, from http://www.hupx.hu/market/electricity/auction_aggregated_curve.html
26. Hungarian Power Exchange (HUPX). (2014b). *Market coupling.* Retrieved May 7, 2014, from http://www.hupx.hu/market-coupling/faq/index.html
27. Hungarian Power Exchange (HUPX). (2014c). *Trading at HUPX.* Retrieved May 5, 2014, from http://www.hupx.hu/trading/index.html
28. International Energy Association (IEA). (2011). *Energy policies of IEA countries, Hungary 2011 review.* Retrieved May 20, 2014, from http://www.iea.org/publications/freepublications/publication/name,34632,en.html
29. Keles, D., Genoese, M., Möst, D., Ortlieb, S., & Fichtner, W. (2013). A combined modeling approach for wind power feed-in and electricity spot prices. *Energy Policy,*

59, 213–225. Retrieved May 12, 2014, from http://www.sciencedirect.com/science/article/pii/S0301421513001821

30. Kolcun, M., Oleinkova, I., & Truicki, M. (2012). Statistically-based analysis of the electricity market price drivers. *Energetika.* Retrieved May 5, 2014, from http://www.lmaleidykla.lt/ojs/index.php/energetika/article/view/2463
31. Krätzig, M., & Lütkepohl, H. (2004). *Applied time series econometrics*. Cambridge: Cambridge University Press.
32. Li, X. R., Yu, C. W., Ren, S. Y., Chiu, C. H., & Meng, K. (2013). Day-ahead electricity price forecasting based on panel cointegration and particle filter. *Electric Power Systems Research, 95*, 66–76. Retrieved May 5, 2014, from http://www.sciencedirect.com/science/article/pii/S0378779612002453
33. Linear least squares. (n.d.) In *Wikipedia.* Retrieved May 31, 2014, from http://en.wikipedia.org/wiki/Linear_least_squares_(mathematics)
34. Liu, H., & Shi, J. (2013). Applying ARMA–GARCH approaches to forecasting short-term electricity prices. *Energy Economics, 37,* 152–166. Retrieved May 5, 2014, from http://www.sciencedirect.com/science/article/pii/S0140988313000303
35. Mavir. (2013). *Data of the Hungarian electricity system 2013.* Retrieved May 19, 2014, from http://www.mavir.hu/documents/10262/160379/MAVIR_adatok_20140514_lapozgato.pdf/87a528ec-f959-4892-8aec-c85a37dd19cc
36. Mavir. (2014). *Data publication (wind generation data).* Retrieved May 9, 2014, from http://www.mavir.hu/web/mavir-en/wind-production
37. Misiorek, M., & Weron, R. (2005). *Forecasting spot electricity prices with time series models.* Retrieved May 8, 2014, from http://prac.im.pwr.wroc.pl/~hugo/publ/RWeronMisiorek05_EEM05_z_logo.pdf
38. Murray, B. (2009). *Power market and economics*. West Sussex: John Willey & Sons Ltd.
39. Ockenfels, A., Grimm, V., & Zoettl, G. (2008). *Electricity market design.* Retrieved May 6, 2014, from http://cdn.eex.com/document/38615/gutachten_eex_ockenfels_e.pdf
40. Reka Drava [River Drava]. (n.d.) In *Wikipedia.* Retrieved May 26, 2014, from http://sl.wikipedia.org/wiki/Drava
41. Renewablefacts. (2012). *Renewable facts by country.* Retrieved May 21, 2014, from http://www.renewablefacts.com/country
42. Stata. (2014). *TS-Arima stata guidelines.* Retrieved June 12, 2014, from http://www.stata.com/bookstore/stata12/pdf/ts_arima.pdf
43. Statkraft. (2014). *Romania.* Retrieved May 7, 2014, from http://www.statkraft.com/about-statkraft/ownership-positions/romania/
44. Stoft, S. (2002). *Power system economics*. New York: A John Willey &Sons, Inc.
45. Swider, D. J., & Weber, C. (2007). Extended ARMA models for estimating price developments on day-ahead electricity markets. *Electric Power Systems Research, 77*(5-6), 583–593. Retrieved May 5, 2014, from http://www.sciencedirect.com/science/article/pii/S0378779606001313

46. Tennet. (2014). *Congestion management.* Retrieved May 7, 2014, from http://www.tennet.eu/de/en/customers/congestion-management.html
47. Transelectrica. (2014). *Transparency (wind generation data).* Retrieved May 9, 2014, from https://www.transelectrica.ro/en/web/tel/productie
48. Tsay, R. S. (2005). *Analysis of finacial time series*. New Jersey: John Wiley & Sons, Inc.
49. Vehviläinen, I., & Pyykkönen, T. (2004). Stochastic factor model for electricity spot price - the case of the Nordic market. Retrieved June 6, 2014, from http://lib.tkk.fi/Diss/2004/isbn9512274027/article1.pdf?origin=publicationDetail
50. Wagner, A. (2014). Residual demand modeling and application to electricity pricing. Retrieved May 7, 2014, from http://www.iaeeu2012.it/pdf/Wagnerppt.pdf
51. Woo, C. K., Horowitz, I., Moore, J., & Pacheco, A. (2011). The impact of wind generation on the electricity spot-market price level and variance: The Texas experience. *Energy Policy, 39*(7), 3939–3944. Retrieved June 6, 2014, from http://www.sciencedirect.com/science/article/pii/S0301421511002813
52. World Nuclear Association (WNA). (2014). Nuclear power in Ukraine. *World Nuclear Association.* Retrieved May 20, 2014, from http://world-nuclear.org/info/Country-Profiles/Countries-T-Z/Ukraine/
53. Wunderground. (2014). *Historical weather (temperature data).* Retrieved May 9, 2014, from http://www.wunderground.com/history/
54. Würzburg, K., Labandeira, X., & Linares, P. (2013). Renewable generation and electricity prices: Taking stock and new evidence for Germany and Austria. *Energy Economics, 40*(S1), S159–S171. Retrieved May 5, 2014, from http://www.sciencedirect.com/science/article/pii/S0140988313002065
55. Yan, X., & Chowdhury, N. A. (2013). Mid-term electricity market clearing price forecasting: A hybrid LSSVM and ARMAX approach. *International Journal of Electrical Power & Energy Systems*, *53*, 20–26. Retrieved May 5, 2014, from http://www.sciencedirect.com/science/article/pii/S0142061513001658
56. Zareipour, H. (2006). Price forecasting and optimal operation of wholesale customers in a competitive electricity market. Retrieved May 6, 2014, from https://uwspace.uwaterloo.ca/bitstream/handle/10012/2611/Hamid_Zareipour_Thesis.pdf;jsessionid=AA3A685584A2B8D7A4FB9164A3287162?sequence=1

APPENDIXES

TABEL OF APPENDIXES

Appendix 1: Acronyms

AR	Autoregressive Model
ARIMA	Autoregressive Integrated Moving Average Model
ARIMAX	Autoregressive Integrated Moving Average Model with Exogenous Variables
ARMA	Autoregressive Moving Average Model
ARMAX	Autoregressive Moving Average Model with Exogenous Variables
HUPX	Hungarian Power Exchange
HUPXDAM	Hungarian Power Exhange Day-Ahead Market
MA	Moving Average Model
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MO	Market Operator
MWh	Megawatt Hour
OLS	Ordinary Least Squares
PX	Power Exchange
TSO	Transmission System Operator
Q	Quarter

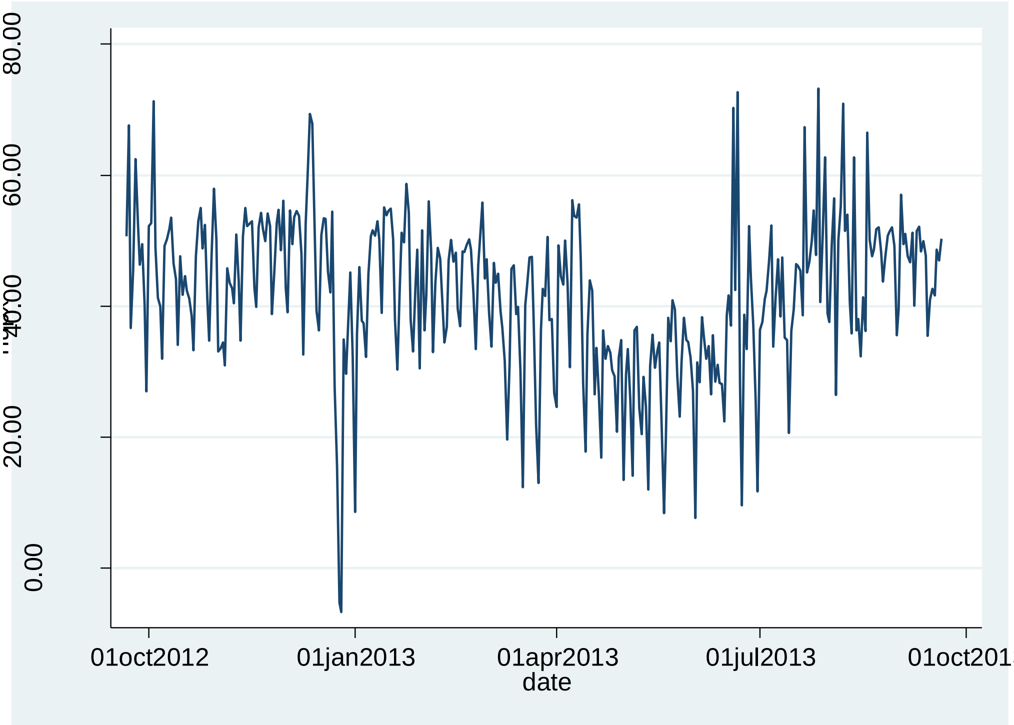
Definitions of all the occurring variables in the following STATA outputs can be found in *Table 4.*

Appendix 2: Temperature Correlation Matrix

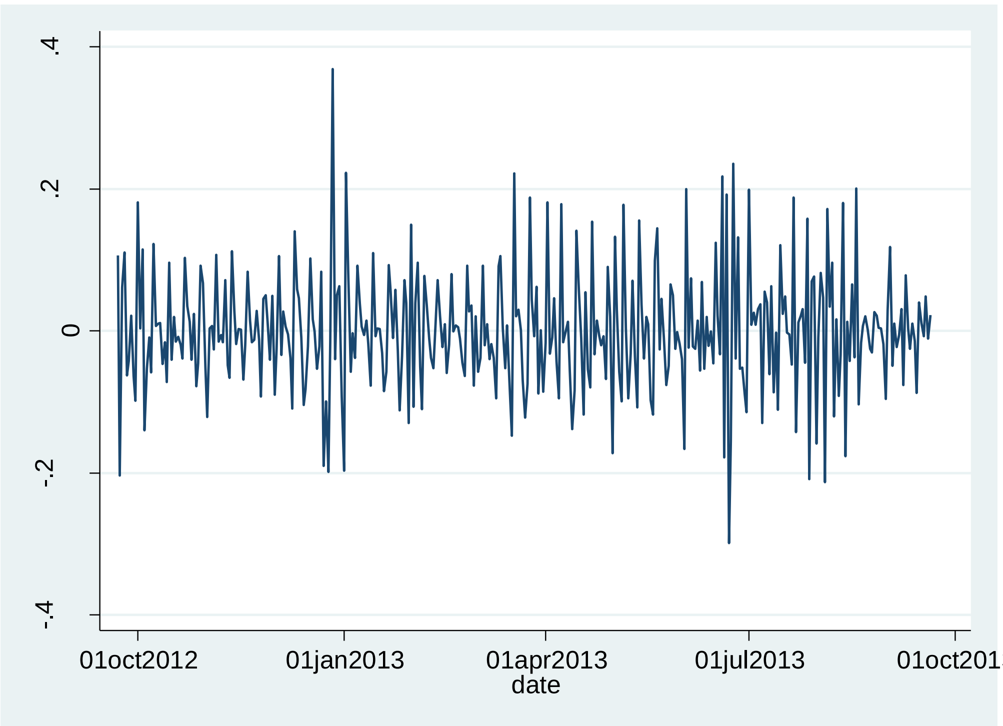
	w_t	t_bu	t_lj	t_sa	t_pg	t_bg	t_zg
w_t	1.0000						
t_bu	0.9728	1.0000					
t_lj	0.9611	0.9110	1.0000				
t_sa	0.9798	0.9271	0.9386	1.0000			
t_pg	0.9629	0.9336	0.9433	0.9480	1.0000		
t_bg	0.9864	0.9313	0.9442	0.9735	0.9391	1.0000	
t_zg	0.9766	0.9147	0.9496	0.9821	0.9399	0.9721	1.0000

*w_t=weighted average temperature, t_x=temperature_x; x=Budapest, Ljubljana, Sarajevo, Podgorica, Belgrade, Zagreb.

Appendix 3: Time-series Line HUPX Spot Price



Appendix 4: Time-series Line HUPX Spot Price (First Difference, Logarithmic Transformation)



Appendix 5: Dickey Fuller Test

Augmented Dickey-Fuller test for unit root   Number of obs = 362

	Test Statistic	Interpolated Dickey-Fuller 1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-19.120	-3.451	-2.876	-2.570

MacKinnon approximate p-value for Z(t) = 0.0000

D.dlnhupx	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
dlnhupx						
L1.	-1.545921	.0808525	-19.12	0.000	-1.704925	-1.386918
LD.	.2325218	.0508677	4.57	0.000	.1324857	.3325579
_cons	.0001398	.0041599	0.03	0.973	-.0080412	.0083207

Appendix 6: Variance Inflationary Table (VIF)

Variable	VIF	1/VIF
hol_cro	1.91	0.523544
hol_slo	1.89	0.528892
hol_hu	1.62	0.618211
hol_ro	1.42	0.706056
ln_w_t	1.21	0.828225
ln_hydro	1.15	0.871233
hol_sr	1.14	0.876733
ln_eol_ro	1.11	0.901790
ln_eol_hu	1.07	0.932990
d_sun	1.07	0.935271
d_sat	1.04	0.959878
Mean VIF	1.33	

Appendix 7: ARIMAX Full Time-span

```
ARIMA regression

Sample:  22sep2012 - 20sep2013                  Number of obs      =       364
                                                Wald chi2(14)      =    948.93
Log pseudolikelihood =  526.0271                Prob > chi2        =    0.0000
```

D.ln_hupx	Coef.	Semirobust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ln_hupx						
d_sat						
D1.	-.0410726	.0085943	-4.78	0.000	-.0579171	-.024228
d_sun						
D1.	-.0926663	.0104262	-8.89	0.000	-.1131012	-.0722313
hol_hu						
D1.	-.0620344	.0227895	-2.72	0.006	-.1067011	-.0173678
hol_cro						
D1.	-.055517	.0234113	-2.37	0.018	-.1014024	-.0096317
hol_slo						
D1.	-.038014	.0158516	-2.40	0.016	-.0690827	-.0069454
hol_ro						
D1.	-.0381755	.0147801	-2.58	0.010	-.0671439	-.0092071
hol_sr						
D1.	-.0070856	.015729	-0.45	0.652	-.037914	.0237427
ln_w_t						
D1.	-.0368856	.1247708	-0.30	0.768	-.2814319	.2076606
ln_eol_hu						
D1.	.0069442	.0088972	0.78	0.435	-.0104939	.0243824
ln_eol_ro						
D1.	.0088569	.0059714	1.48	0.138	-.0028469	.0205606
ln_hydro						
D1.	-.0250792	.0580341	-0.43	0.666	-.1388239	.0886656
summer						
D1.	.0829771	.1136911	0.73	0.465	-.1398535	.3058076
_cons	-.0003145	.0004007	-0.78	0.432	-.0010998	.0004708
ARMA						
ar						
L1.	.35612	.2226187	1.60	0.110	-.0802047	.7924446
ma						
L1.	-.9535198	.1922195	-4.96	0.000	-1.330263	-.5767766
/sigma	.0569018	.0034079	16.70	0.000	.0502225	.0635811

Appendix 8: Q1 Summary

Variable	Obs	Mean	Std. Dev.	Min	Max
hol_hu	90	.0222222	.1482314	0	1
hol_cro	90	.0222222	.1482314	0	1
hol_slo	90	.0222222	.1482314	0	1
hol_ro	90	.0222222	.1482314	0	1
hol_sr	90	.0222222	.1482314	0	1
hupx	90	47.54125	8.832983	27.03167	71.32792
hydro	90	439.6482	180.4237	234.021	1416.979
eol_hu	90	77.55222	64.89508	1.484958	257.3961
eol_ro	90	394.1733	305.0045	8.169542	1369.001
w_t	90	9.218137	6.937702	-7.797647	21.85428
hupx	90	47.54125	8.832983	27.03167	71.32792

Appendix 9: Q2 Summary

Variable	Obs	Mean	Std. Dev.	Min	Max
hol_hu	90	.0444444	.2072349	0	1
hol_cro	90	.0444444	.2072349	0	1
hol_slo	90	.0555556	.2303447	0	1
hol_ro	90	.0666667	.2508413	0	1
hol_sr	90	.0777778	.269322	0	1
hupx	90	41.00686	11.96539	-6.710417	58.71292
hydro	90	212.1114	44.69507	122	348
eol_hu	90	92.44342	73.22519	1.519375	287.0095
eol_ro	90	581.7038	369.1413	14.0345	1422.701
w_t	90	2.388193	3.456714	-3.775069	11.38989
hupx	90	41.00686	11.96539	-6.710417	58.71292

Appendix 10: Q3 Summary

Variable	Obs	Mean	Std. Dev.	Min	Max
hol_hu	90	.0555556	.2303447	0	1
hol_cro	90	.0333333	.1805111	0	1
hol_slo	90	.0555556	.2303447	0	1
hol_ro	90	.0444444	.2072349	0	1
hol_sr	90	.0777778	.269322	0	1
hupx	90	32.8141	10.32123	7.605417	56.17625
hydro	90	439.1111	113.5078	229	642
eol_hu	90	76.6931	63.9442	2.383458	275.3661
eol_ro	90	532.9523	355.8259	42.63458	1568.657
w_t	90	14.11959	5.537133	-.3786158	23.83216
hupx	90	32.8141	10.32123	7.605417	56.17625

Appendix 11: Q4 Summary

Variable	Obs	Mean	Std. Dev.	Min	Max
hol_hu	95	.0105263	.1025978	0	1
hol_cro	95	.0315789	.175804	0	1
hol_slo	95	.0210526	.1443214	0	1
hol_ro	95	.0315789	.175804	0	1
hol_sr	95	.0105263	.1025978	0	1
hupx	95	45.23156	11.1549	9.561667	73.19167
hydro	95	278.4316	94.3409	161	555
eol_hu	95	58.44925	53.88238	.6124583	219.56
eol_ro	95	386.832	295.2831	43.85046	1599.719
w_t	95	21.16353	3.201936	12.38977	27.63411
hupx	95	45.23156	11.1549	9.561667	73.19167

Appendix 12: Q1 ARIMA

ARIMA regression

Sample: 22sep2012 - 19dec2012 — Number of obs = 89

Wald chi2(2) = 5.26e+10

Log pseudolikelihood = 129.0038 — Prob > chi2 = 0.0000

D.ln_hupx	Coef.	Semirobust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ln_hupx						
_cons	.0003547	.00032	1.11	0.268	-.0002725	.0009819
ARMA						
ar						
L1.	.3552788	.1084461	3.28	0.001	.1427283	.5678293
ma						
L1.	-1.000003	4.41e-06	-2.3e+05	0.000	-1.000012	-.9999943
/sigma	.0555984	.0038618	14.40	0.000	.0480293	.0631674

Appendix 13: Q1 ARIMAX, with All Significant Exogenous Variables in Q1, and Used for Prediction

```
ARIMA regression

Sample:  22sep2012 - 19dec2012                  Number of obs      =        89
                                                Wald chi2(5)       =    624.64
Log pseudolikelihood =  162.0273                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
             |             Semirobust
  D.ln_hupx  |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ln_hupx      |
      ln_w_t |
         D1. |  -.4179705    .1766904    -2.37   0.018    -.7642773   -.0716637
             |
       d_sat |
         D1. |  -.0382775    .0117756    -3.25   0.001    -.0613573   -.0151978
             |
       d_sun |
         D1. |  -.1068764    .0102132   -10.46   0.000     -.126894   -.0868588
             |
       _cons |   -.000762    .0007675    -0.99   0.321    -.0022663    .0007422
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |   .2778278    .1627864     1.71   0.088    -.0412276    .5968832
             |
          ma |
         L1. |  -1.109368    .0845555   -13.12   0.000    -1.275094   -.9436424
-------------+----------------------------------------------------------------
      /sigma |   .0350887    .0046351     7.57   0.000     .0260041    .0441733
------------------------------------------------------------------------------
```

Appendix 14: Q2 ARIMA

```
ARIMA regression

Sample:  20dec2012 - 19mar2013                  Number of obs      =         90
                                                Wald chi2(2)       =   6.03e+11
Log pseudolikelihood =  101.6657                Prob > chi2        =     0.0000
```

D.ln_hupx	Coef.	Semirobust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ln_hupx						
_cons	.0002924	.0005946	0.49	0.623	-.000873	.0014579
ARMA						
ar						
L1.	.573875	.1466322	3.91	0.000	.2864812	.8612687
ma						
L1.	-.999999	1.41e-06	-7.1e+05	0.000	-1.000002	-.9999963
/sigma	.0767996	.0074956	10.25	0.000	.0621085	.0914908

Appendix 15: Q2 ARIMAX, with All Significant Exogenous Variables in Q2, AR Part Insignificant

```
ARIMA regression

Sample:  20dec2012 - 19mar2013                  Number of obs      =        90
                                                Wald chi2(7)       =    250.20
Log pseudolikelihood =  136.1572                Prob > chi2        =    0.0000
```

D.ln_hupx	Coef.	Semirobust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ln_hupx						
ln_hydro						
D1.	.3214793	.0863408	3.72	0.000	.1522545	.4907041
eol_ro						
D1.	.0000309	.0000187	1.65	0.098	-5.73e-06	.0000674
hol_cro						
D1.	-.0562045	.0229781	-2.45	0.014	-.1012407	-.0111683
hol_slo						
D1.	-.0394115	.0212254	-1.86	0.063	-.0810125	.0021894
new_year						
D1.	-.1522087	.0420781	-3.62	0.000	-.2346802	-.0697373
_cons	-.000674	.0022715	-0.30	0.767	-.0051261	.003778
ARMA						
ar						
L1.	.4471707	.3104525	1.44	0.150	-.161305	1.055646
ma						
L1.	-.7799258	.1751643	-4.45	0.000	-1.123242	-.43661
/sigma	.053211	.0048487	10.97	0.000	.0437078	.0627143

Appendix 16: Correlation Matrix for Q2 (Spot Price, Hydro Production and Wind Production in Romania)

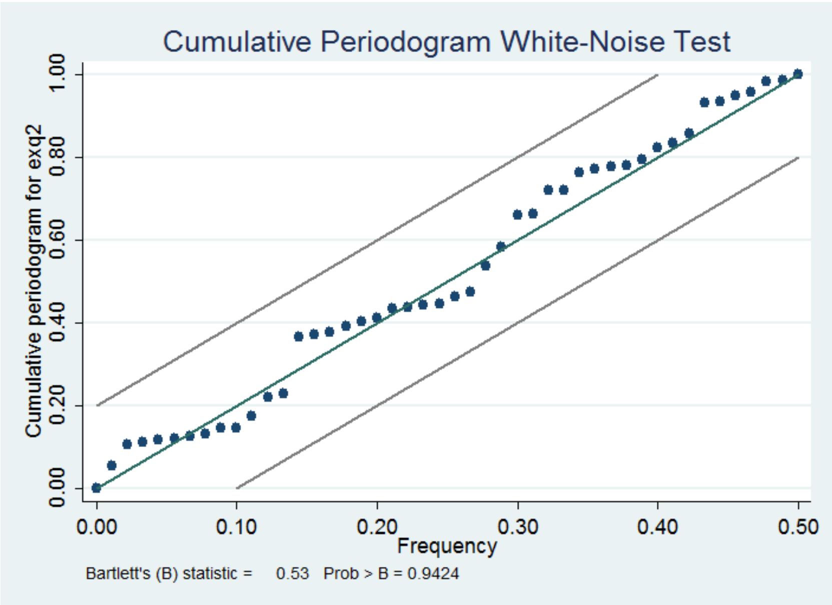
	ln_hupx	ln_hydro	ln_eol~o
ln_hupx	1.0000		
ln_hydro	-0.0745	1.0000	
ln_eol_ro	0.3036	-0.0091	1.0000

Appendix 17: Q2 ARIMAX with Used Variables for Prediction, AR Part Significant

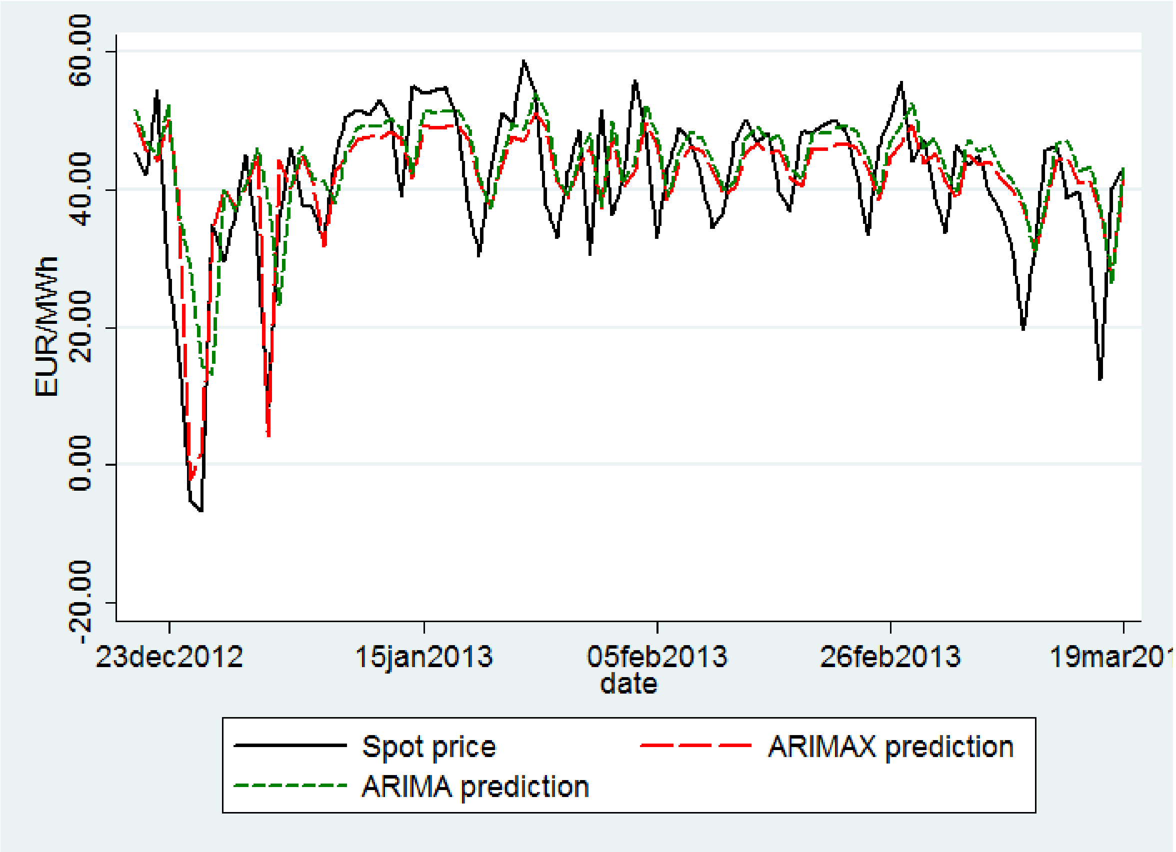
```
ARIMA regression

Sample:  20dec2012 - 19mar2013                  Number of obs      =        90
                                                Wald chi2(4)       =   3.78e+08
Log pseudolikelihood =  124.7289                Prob > chi2        =     0.0000
```

D.ln_hupx	Coef.	Semirobust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ln_hupx						
hol_cro						
D1.	-.067635	.0286216	-2.36	0.018	-.1237323	-.0115376
new_year						
D1.	-.2160711	.0457185	-4.73	0.000	-.3056777	-.1264645
_cons	-.0001696	.0004593	-0.37	0.712	-.0010698	.0007305
ARMA						
ar						
L1.	.4813242	.1315661	3.66	0.000	.2234593	.739189
ma						
L1.	-.9999785	.0000517	-1.9e+04	0.000	-1.00008	-.9998772
/sigma	.0593605	.0049884	11.90	0.000	.0495834	.0691375

Appendix 18: Q2 Test for White Noise Residuals



Appendix 19: Predicted and Actual Spot Price for Q2 (In-sample Prediction)



Appendix 20: MAE and MAPE for Q2

Error measurement	ARIMA	ARIMAX
MAE(€/MWh)	7,69	6,49
MAPE (%)	34,65	21,25

Appendix 21: Q3 ARIMA

ARIMA regression

Sample: 20mar2013 - 17jun2013	Number of obs	=	90
	Wald chi2(2)	=	1.19e+12
Log pseudolikelihood = 111.3813	Prob > chi2	=	0.0000

D.ln_hupx	Coef.	Semirobust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ln_hupx						
_cons	-.0008983	.0004116	-2.18	0.029	-.001705	-.0000916
ARMA						
ar						
L1.	.4013639	.1003334	4.00	0.000	.2047141	.5980137
ma						
L1.	-1	9.63e-07	-1.0e+06	0.000	-1.000002	-.9999982
/sigma	.0687743	.0050735	13.56	0.000	.0588303	.0787182

Appendix 22: Q3 ARIMAX with All Significant Variables

```
ARIMA regression

Sample:  20mar2013 - 17jun2013                  Number of obs      =         90
                                                Wald chi2(7)       =   1.90e+11
Log pseudolikelihood =  155.2595                Prob > chi2        =     0.0000
```

D.ln_hupx	Coef.	Semirobust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ln_hupx						
ln_hydro						
D1.	-.20568	.0607893	-3.38	0.001	-.3248248	-.0865352
ln_w_t						
D1.	.6698455	.2790436	2.40	0.016	.12293	1.216761
d_sat						
D1.	-.0582848	.0114806	-5.08	0.000	-.0807864	-.0357832
d_sun						
D1.	-.1178321	.0150306	-7.84	0.000	-.1472914	-.0883727
hol_hu						
D1.	-.120309	.0312683	-3.85	0.000	-.1815939	-.0590242
_cons	-.0007243	.0003906	-1.85	0.064	-.0014898	.0000411
ARMA						
ar						
L1.	.3672563	.1157797	3.17	0.002	.1403322	.5941804
ma						
L1.	-1.000004	3.01e-06	-3.3e+05	0.000	-1.00001	-.9999985
/sigma	.042218	.0037166	11.36	0.000	.0349336	.0495024

Appendix 23: Q3 ARIMAX Used for Prediction

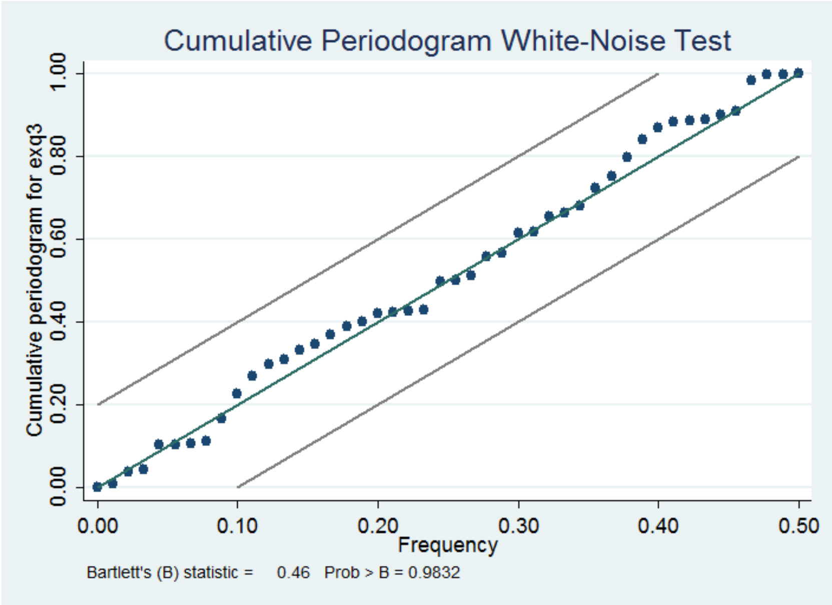
```
ARIMA regression

Sample:  20mar2013 - 17jun2013                  Number of obs      =        90
                                                Wald chi2(6)       =  1.69e+14
Log pseudolikelihood =   152.311                Prob > chi2        =    0.0000
```

D.ln_hupx	Coef.	Semirobust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ln_hupx						
ln_hydro						
D1.	-.1259814	.0444611	-2.83	0.005	-.2131235	-.0388393
d_sat						
D1.	-.0547675	.0120787	-4.53	0.000	-.0784413	-.0310937
d_sun						
D1.	-.1154715	.0162271	-7.12	0.000	-.147276	-.083667
hol_hu						
D1.	-.1132429	.0323806	-3.50	0.000	-.1767077	-.049778
_cons	-.0001916	.0004193	-0.46	0.648	-.0010135	.0006303
ARMA						
ar						
L1.	.4093034	.1025869	3.99	0.000	.2082368	.61037
ma						
L1.	-1	9.82e-08	-1.0e+07	0.000	-1	-.9999999
/sigma	.043648	.0043498	10.03	0.000	.0351225	.0521735

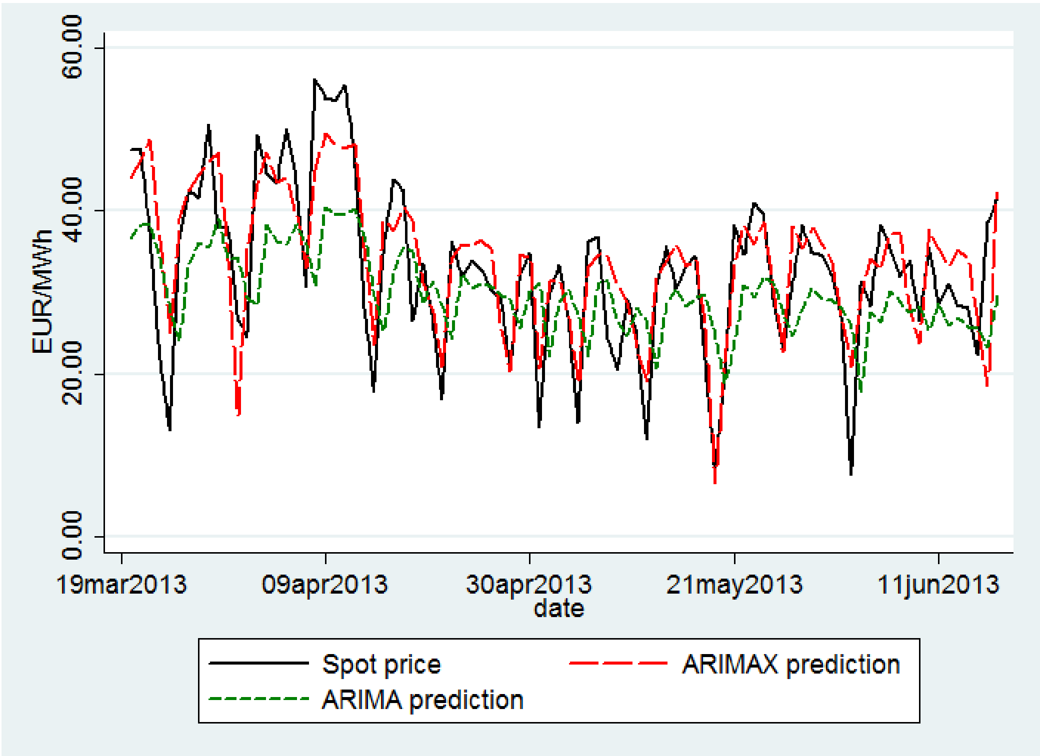
Appendix 24: Correlation Matrix for Q3 (Spot price, Hydro Production and Temperature)

	ln_hupx	ln_hydro	ln_w_t
ln_hupx	1.0000		
ln_hydro	-0.4546	1.0000	
ln_w_t	-0.3518	0.8371	1.0000

Appendix 25: Q3 Test For White Noise Residuals



Appendix 26: Predicted and Actual Spot Price for Q3 (In-sample Prediction)



Appendix 27: MAE and MAPE for Q3.

Error measurement	ARIMA	ARIMAX
MAE(€/MWh)	7,76	4,41
MAPE (%)	29,49	17,22

Appendix 28: Q4 ARIMA, AR Part Insignificant

```
ARIMA regression

Sample:  19jun2013 - 20sep2013                  Number of obs     =         94
                                                Wald chi2(2)      =   1.35e+10
Log pseudolikelihood =  108.3555                Prob > chi2       =     0.0000
```

D.ln_hupx	Coef.	Semirobust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ln_hupx						
_cons	.0007144	.000239	2.99	0.003	.0002461	.0011828
ARMA						
ar						
L1.	.1697175	.1223465	1.39	0.165	-.0700773	.4095123
ma						
L1.	-1.000006	8.75e-06	-1.1e+05	0.000	-1.000024	-.9999893
/sigma	.0747145	.0072863	10.25	0.000	.0604337	.0889953

Appendix 29: Q4 ARIMA (1,1,0)

```
ARIMA regression

Sample:  19jun2013 - 20sep2013                  Number of obs     =         94
                                                Wald chi2(1)      =      10.94
Log pseudolikelihood =  95.96735                Prob > chi2       =     0.0009
```

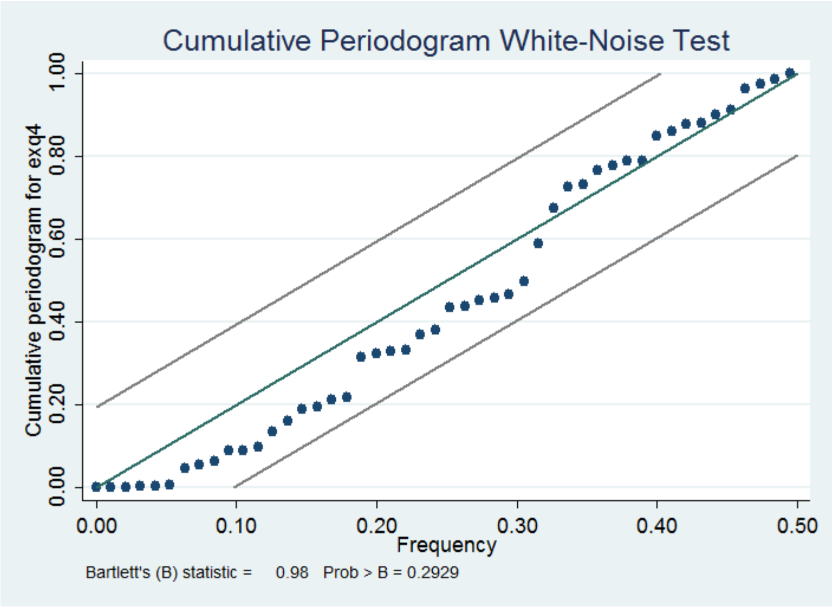
D.ln_hupx	Coef.	Semirobust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ln_hupx						
_cons	.0001658	.0060528	0.03	0.978	-.0116975	.012029
ARMA						
ar						
L1.	-.4685651	.1416661	-3.31	0.001	-.7462255	-.1909047
/sigma	.0870579	.0074712	11.65	0.000	.0724147	.1017012

Appendix 30: Q4 ARIMAX, with All Significant Exogenous Variables in Q4 and Used for Prediction

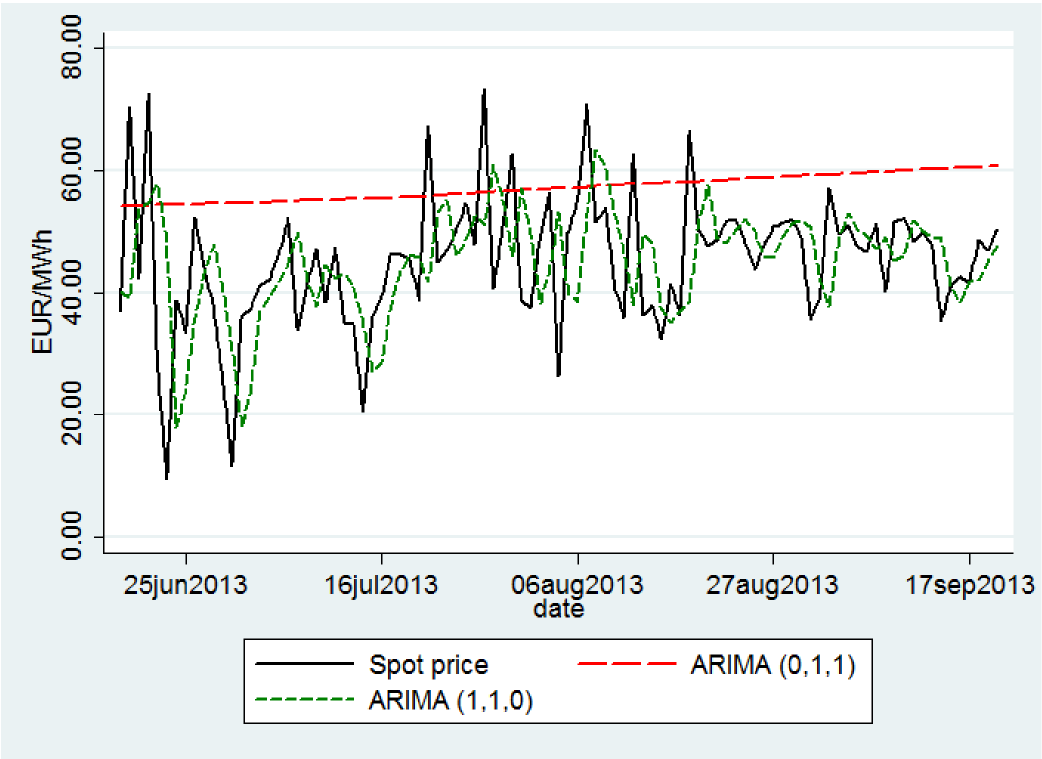
```
ARIMA regression

Sample:  19jun2013 - 20sep2013                  Number of obs      =        94
                                                Wald chi2(7)       =     68.26
Log pseudolikelihood =  113.6666                Prob > chi2        =    0.0000
```

D.ln_hupx	Coef.	Semirobust Std. Err.	z	P>|z|	[95% Conf.	Interval]
ln_hupx						
ln_w_t						
D1.	.4710514	.6215347	0.76	0.449	-.7471343	1.689237
d_sat						
D1.	-.0402327	.0192642	-2.09	0.037	-.0779898	-.0024756
d_sun						
D1.	-.0731568	.0220004	-3.33	0.001	-.1162768	-.0300369
hol_hu						
D1.	-.0927657	.0290389	-3.19	0.001	-.1496809	-.0358505
hol_cro						
D1.	-.1251845	.0445837	-2.81	0.005	-.212567	-.037802
hol_ro						
D1.	-.0866357	.0310966	-2.79	0.005	-.1475839	-.0256875
_cons	.0004968	.0047	0.11	0.916	-.0087149	.0097086
ARMA						
ar						
L1.	-.6284994	.1038245	-6.05	0.000	-.8319916	-.4250072
/sigma	.0720189	.0063157	11.40	0.000	.0596404	.0843975

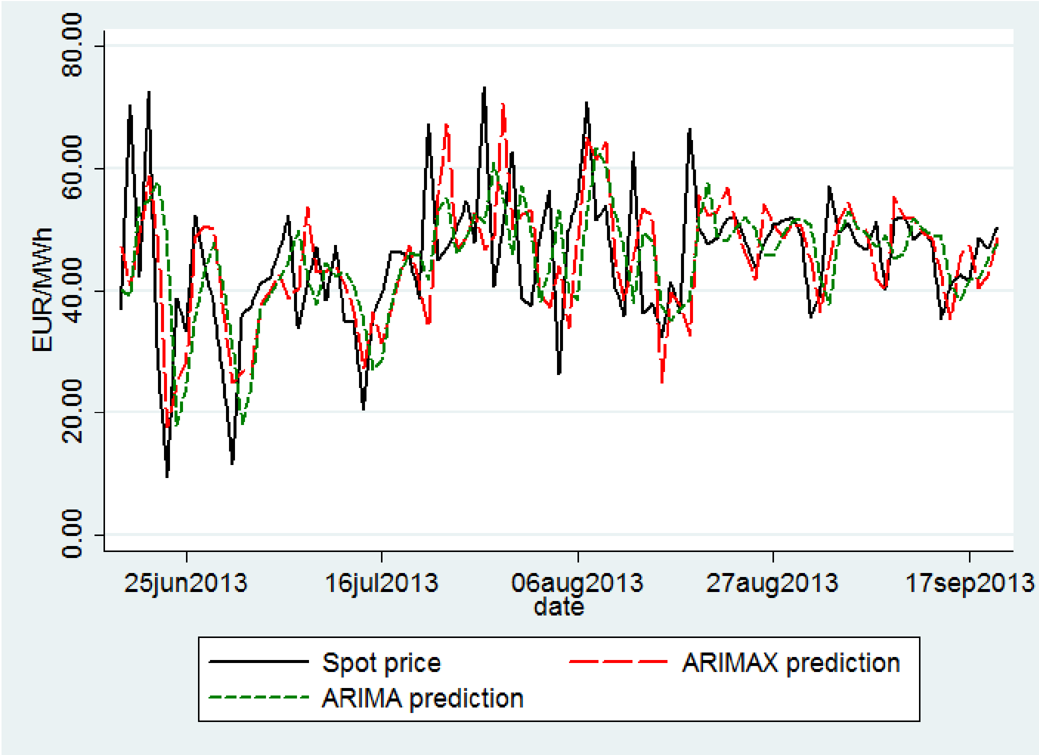
Appendix 31: Q4 Test for White Noise Residuals



Appendix 32: Comparison of Prediction by Different ARIMA Models for Q4 (In-sample Prediction)



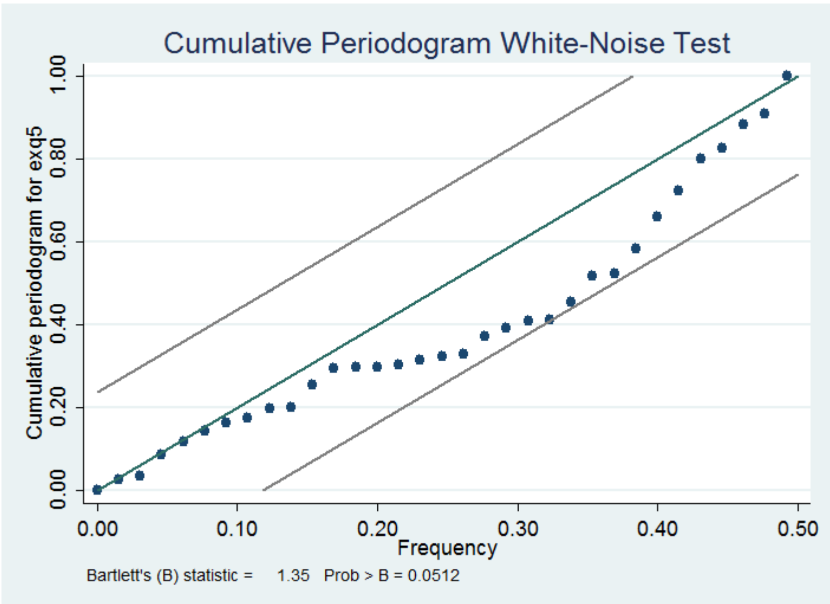
Appendix 33: Predicted and Actual Spot Price for Q4 (In-sample Prediction)



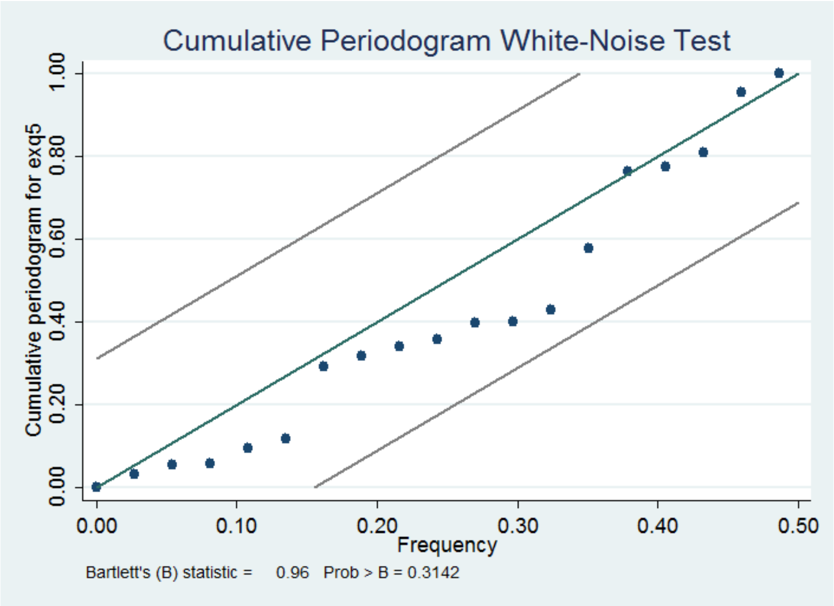
Appendix 34: MAE and MAPE for Q4

Error measurement	ARIMA	ARIMAX
MAE(€/MWh)	9,45	8,04
MAPE (%)	26,04	19,42

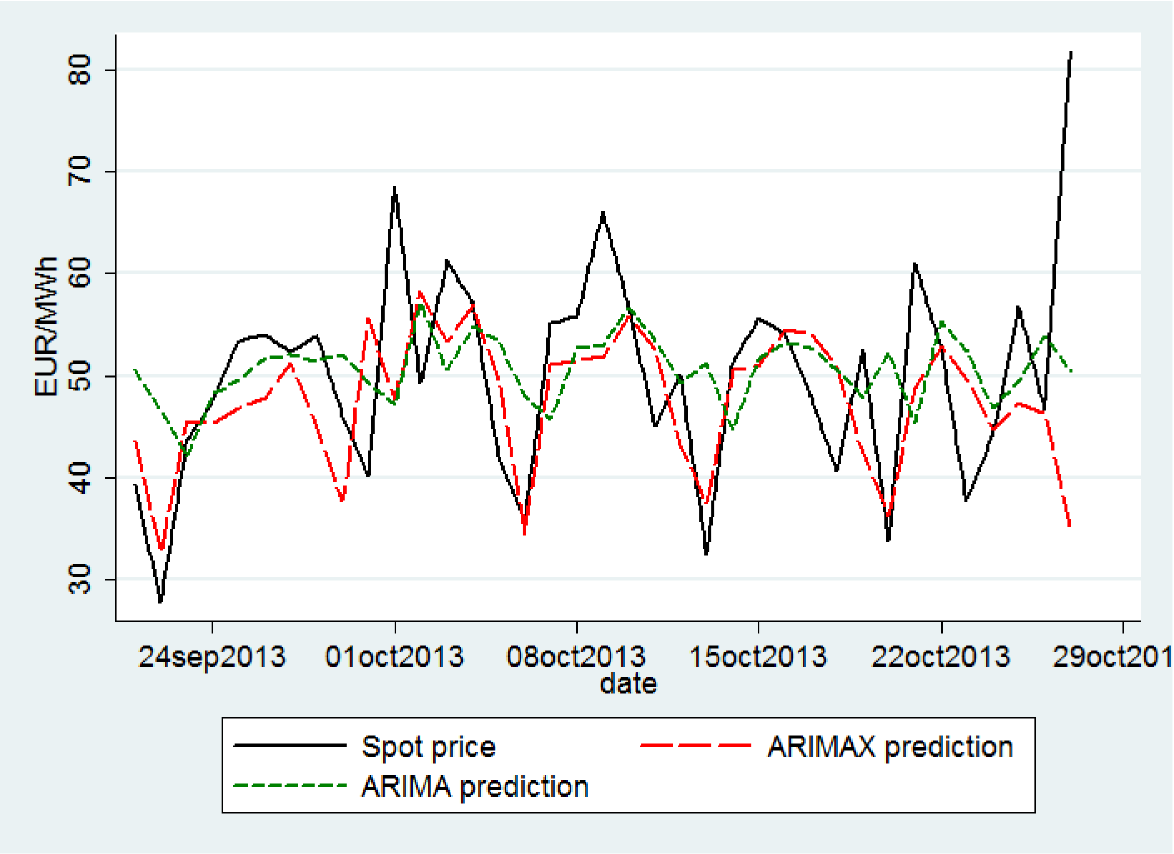
Appendix 35: Q5 Test for White Noise Residuals, Full Time Span



Appendix 36: Q5 Test for White Noise Residuals, until 27 October 2013



Appendix 37: Forecasted and Actual Spot Price for Q5 until 27 October 2013 (Out-of-sample Forecast)



Appendix 38: MAE and MAPE for Q5

Error measurement	ARIMA	ARIMAX
MAE(€/MWh)	8,28	7,17
MAPE (%)	17,98	13,68