UNIVERSITY OF BASQUE COUNTRY

MASTER IN ECONOMICS: EAP

Master Thesis

Electricity Price Forecasting using Sale and Purchase Curves: The X-Model

Evidence for Spain

eman ta zabal zazu

Universidad del País Vasco
Euskal Herriko Unibertsitatea

Author:

Jorge Rey Colorado

Supervisors:

Cristina Pizarro-Irizar and Peru Muniain Izagirre

July 29, 2022

Abstract

In this Master Thesis, we implement in the Spanish electricity market the X-Model proposed by Ziel and Steinert (2016), combining the insights of market structure models with econometric techniques. Instead of directly modeling the electricity price using past electricity prices, we model and utilize its true source: the sale and purchase curves of the electricity market. We incorporate not only several known features, such as seasonal behavior or renewable energy participation, but also other facts of the bidding structure, such as price clustering. The model is able to capture the non-linear behavior of the electricity price, which is especially useful for predicting huge price spikes. We simulate the months of April 2019 and 2020 and we observe that the model performs better for demand than for supply. We also observe that the supply curve forecast is sensitive to price spikes and that the fit improves as the renewable participation increases. From the policy perspective, this Master Thesis can help to identify the effects of supply-induced changes (such as fossil fuel price changes, the phase-out of coal or nuclear plants or the integration of more renewables) and demand-induced changes (such as the effect of the Covid-19 pandemic). All these issues are in the current energy debate on energy economics.

Keywords: bid structure, electricity market, electricity modelling, electricity prices.

Contents

List of Figures

1 Introduction

The integrated Iberian Electricity Market (MIBEL) started operating on the 1st of July 2007 and is formed by the integration of the Portuguese and Spanish markets. Therefore, it is expected to obtain the same price for both regions in the equilibrium of the day-ahead market, if there were no transmission congestion. Also, the good interchanged in this market is electricity, which is homogeneous, strategic, and non-storable.

In the mid-1980, many electricity markets started liberalization reforms which transformed the electricity sector from a vertically-integrated monopoly to a competitive market. The liberalization of the electricity sector of the European Union was launched in 1996 and the Spanish government introduced the Electricity Sector Act (ESA) in 1996 (Ciarreta et al. (2016)), and due to the liberalization of the markets and the increasing disclosure of data, new insights concerning the structure and behavior of the prices were gained (Ziel and Steinert (2016)), thereby, forecasting prices has become an important task for the agents involved.

However, in (Weron (2007)) we found some stylized facts about the energy prices, the most important one is the presence of price spikes. Electricity prices exhibit extreme volatility up to 50%, in addition, The spike intensity is also non-homogeneous in time and is especially notorious during *on-peak hours*, i.e, around 09:00 and 18:00 on business days. This characteristic has a huge impact on researchers as well as politics and companies. Also long-term cost calculation for investment projects or political as the renewable transition (Ziel and Steinert (2016)).

There is a wide variety of models for estimating energy prices, most of them come from or are related to finance literature and can be divided into 5 groups as follows (Weron (2014)): Multi-agent, Fundamental (structural), Reduced-form (Quantitative, stochastic), statistical (econometric) and computational intelligence. However, most price forecasting approaches are hybrid solutions and focus only on the price itself.

In this sense, in this Master Thesis, we are going to forecast the energy prices for Spain using a different approach. In particular, we use "The X model" proposed by Ziel and Steinert (Ziel and Steinert (2016)), this approach is interesting because it uses the actual source, the supply, and demand curves of the electricity market. Therefore, we have more information and we can obtain better results than other models that simply use the prices. In particular, instead of directly modeling the electricity price using past electricity prices, we model and utilize its true source: the sale and purchase curves of the electricity market. We incorporate not only several known features, such as seasonal behavior or renewable energy participation, but also other facts of the bidding structure, such as price clustering. The model is able to capture the non-linear behavior of the electricity price, which is especially useful for predicting huge price spikes. From the policy perspective, this Master Thesis can help to identify the effects of supply-induced changes (such as fossil fuel price changes, the phase-out of coal or nuclear plants or the integration of more renewables) and demand-induced changes (such as the effect of the Covid-19 pandemic). All these issues are in the current energy debate on energy economics.

Therefore, this Master Thesis wants to determine how well the model performs in forecasting the Spanish electricity market prices. We forecast supply and demand curves for the months of April 2019 and 2020, so that we can assess how well the model performs in the presence of economic shocks. In particular, we can observe the demand shock due to Covid-19 and quarantine during that period.

The Master Thesis is structured as follows. *Section 2* will provide a brief literature review on forecasting the electricity prices and why this model is interesting and also a good approach. *Section 3* will explain -in detail- the methodology used and the data. *Section 4* will discuss the empirical results and finally in *Section 5* conclusions on the results obtained will be presented.

2 Literature Review

A great variety of models for forecasting the electricity price can be found, most of the are often related to the well-known model of the finance literature. Weron divided them into five different groups (Weron (2014)), multi-agent, fundamental (structural), reduced-form (Quantitative, stochastic), statistical (econometric), and computational intelligence.

Multi-agent model focus mainly focus on the supply and demand of electricity to obtain prices by equilibrium optimization (Weron (2014); Ziel and Steinert (2016)) for instance, the literature has used production cost models (PCMs), however, these model ignored the oligopolistic structure of the market, therefore, Battle López and Barquil Gil, propose a Strategic Production Costing Model (SPCM) which can model the oligopolistic model and their bidding process (Batlle and Barquin (2005)), nevertheless, these kinds of model do not incorporate the time series of electricity bids, for instance (Ventosa et al. (2005)), furthermore, cannot generally provide accurate daily or hourly, price forecast as can be seen in the attempt made for the Spanish market (García-Alcalde et al. (2002)).

The fundamental or structural models cover a great variety of models but mainly emphasize economic and physical relationships of the market (Carmona and Coulon (2014); Weron (2014)). Moreover, there is an important stylized fact of the electricity prices. There is a tremendous deviation of the price from its mean, up to 50.% (Weron (2007)), therefore, in the area of time series models the usage of specific heteroscedastic models for the variance of the process is used, but, standard GARCH model cannot account for all of the extreme price events within the data (Swider and Weber (2007)) such as (ARMA-GARCH) (Liu and Shi (2013)), nevertheless, all these model foci mainly on the prices time series and not in what determines the process (Ziel and Steinert (2016)).

Most of the models used in the literature belong to the field of fundamental models because the time series approach is relatively new and has not been applied appropriately, therefore, modeling the structure of the supply and demand curve have been done (Ziel and Steinert (2016)). For the Germany/Austria EPEX market the use of the supply and demand curve have been proposed in 2012 (Eichler et al. (2012)), Coulon utilized the curves to model a scaled supply and demand curve using an autoregressive time series model with weekdays effect and tried to overcome the issue of the assumption of the inelastic demand (Coulon et al. (2014)). In Spain, something similar was done, however, they used a “residual demand curve” (Aneiros et al. (2013)), although none of them incorporate real auction data, they assume the demand is inelastic and therefore focus on the bid stack.

3 Methodology

We are going to use the true data generation process, e.g. the supply and demand curve of the energy price, this approach provides a better probabilistic forecast for extreme price movements while modeling the time series of energy prices by an autoregressive approach (Ziel and Steinert (2016)). Hourly day-ahead energy price auction data of Spain provided by the nominated electricity market operator (NEMO[1]) Every price will be provided in EUR/MWh and every volume in MW, if not specified otherwise. Note that the market clearing volume is reported by the OMIE as energy in MWh. In this Master Thesis, we are going to work with simple bids instead of complex bids, due to the lack of time, however for further research and looking for publishing we can consider the possibility. In addition, we have to say that real prices of the electricity market are found by the well-known **EUPHEMIA** algorithm, which we are not going to consider in the model due to its complexity. In addition, we are going to use generation forecast and wind-solar generation forecasts from **ENTSOE**, which is the European association for the cooperation of transmission systems.

3.1 Price formation process and price curves structure

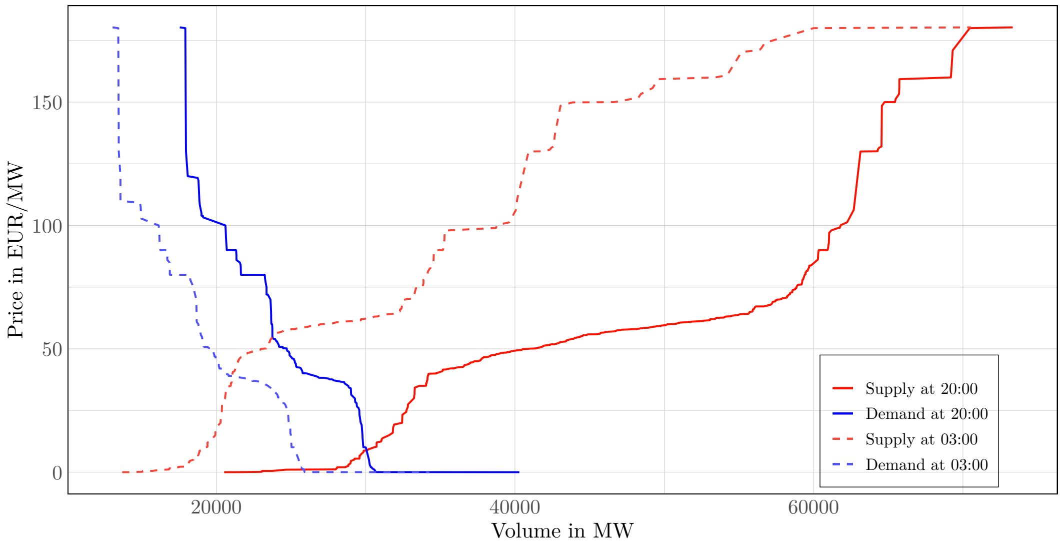
Anyway, some authors point out that the price is driven by external factors, for instance, wind and solar or electricity demand (Weron (2014)). However, Ziel and Steinert (2016) claim that buyers and sellers on an electricity exchange are the ones that are influenced by those factors and therefore adjust their bids. Furthermore, the participants are not equal, not all electricity producers are equal, they have different production portfolios, and thus, are more or less likely affected by heavy weather conditions. All this information is captured in the supply and demand curve.

To show the characteristic behavior of the Spanish electricity market we plot Figure 1, this figure shows the day-ahead electricity supply and demand curve for the peak hour and the non-peak hours. The horizontal axis represents the trading volume and the vertical axis represents the price(bid) associated with the volume, the maximum price(bid) possible is 180.30. It can be observed that during peak hours the amount of electricity produced and demanded is larger than during non-peak hours. In addition, from the supply side, the strategic behavior

[1]Operador de Mercado eléctrico designado (OMIE) in Spanish

of the generator is observed in the middle of the graph, this is associated with the production by hydroelectric power plants which tend to be strategic because of the use of the water as a resource for producing electricity compared with the strategic behavior of solar plants.

Figure 1: Supply and Demand curve at 03:00 and 20:00



The day ahead electricity prices of the OMIE will be traded in daily auctions before 11:00 CET for the hours of the next day. Then, there are in general 24 prices every day. And due to the daylight saving time, there is once a year 23 values in March and 25 values in October. And for the 24 actions on a common day has been used the labels 0:00,0:10, ..., 23:00 throughout this Master Thesis.

The spot price during the period that has been observed is set to be between $P_{min} = 0$ and $P_{max} = 180.30$, we did not take into account 2021 in the analysis, because from may of 2021 the price was set to be between $P_{min} = -500$ and $P_{max} = 3000$ allowing negative bids, due to this year was not include. According to OMIE regulation, the minimal price difference between different orders is 0.01 EUR/MWh. Thus, there are in total 18031 different possible prices on a complete grid $\mathbb{P} = \{0, 0.01, 0.02, 180.28, 180.30\}$, nevertheless, not every price is used on a real auction and yet. Therefore, $\mathbb{P}_{S,t}$ and $\mathbb{P}_{D,t}$ can be introduced, they represent supply and demand bids associated with a positive bid volume.

$$\mathbb{P}_{S,t} = \{P \in \mathbb{P} | V_{S,t}(P) > 0\} \text{ and } \mathbb{P}_{D,t} = \{P \in \mathbb{P} | V_{D,t}(P) > 0\} \tag{1}$$

And if $V_{S,t}, V_{D,t}$ are aggregated for a certain hour the supply and demand curve can be obtained as in Figure 1. Mathematically, the supply and demand curves are characterized by

$$S_t(P) = \sum V_{S,t}(P) \text{for P} \in \text{ and } D_t(P) = \sum V_{D,t}(P) \tag{2}$$

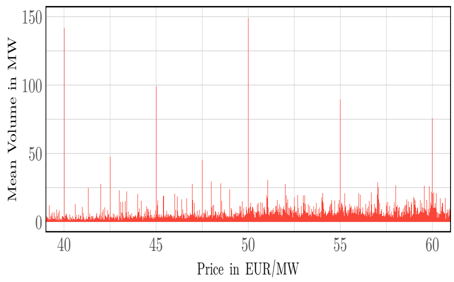
Equation (2) define the supply and demand curve, for the prices grids defined in equation (1). The clearing prices for the Spanish electricity market are defined by the EUPHEMIA algorithm which involves complex bids, however, in this Master Thesis, we are working with simple bids, yet, it is assumed by the MIBEL market as many other markets, i.e, EPEX market (Ziel and Steinert (2016)), that the relation of two different bid price and quantity combinations of one market is linear.

The next step is crucial, time dependency has to be ignored in this first step, thus we can understand the characteristics of the bid volume to subsequently be able to forecast it. Then, the mean bid volume is evaluated for the supply and demand curve across all the databases as in equation (3), and since and since two years period is being considered the number of observations T will be 17520.

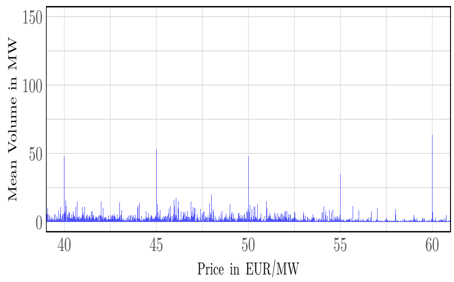
$$\overline{V_S}(P) = \frac{1}{T}\sum_{t=1}^{T} V_{s,t}(P) \; and \; \overline{V_D}(P) = \frac{1}{T}\sum_{t=1}^{T} V_{D,t}(P) \tag{3}$$

In Figure 2 we have plotted the average mean volume for the supply (6a) and the demand (6b) that was obtained using from equation (3). It can be seen in both figures and interesting bidding behavior, operators tend usually bid a higher amount of volume at rounded and multiply of five numbers and we can say that a noticeable amount of tradings is done by human decisions and are not based on algorithmic trading rules which are in line with the theory (Ziel and Steinert (2016)).

Figure 2: Average bid volumes $\overline{V_S}(P)$ and $\overline{V_D}(P)$ for price range 40 to 60



(a) Bid volumes: supply



(b) Bid volumes: Demand

3.2 Model for the supply and demand curve

The model to be used will be the so-called X-model by (Ziel and Steinert (2016)). This model will be used because it has been shown to have a good performance forecasting the supply and demand curve compared to other benchmarks models. As the in the original paper the procedure is divided into three steps:

1. Due to the large number of observations the data has it is mandatory to organize the bid volume in price classes.

2. The stochastic model to forecast the bid volume of each price class is going to be presented next.

3. And once the forecast bids for each price class is obtained we reassemble the bidding structure by reconstructing the classes. Thus, we calculate the supply and demand curve.

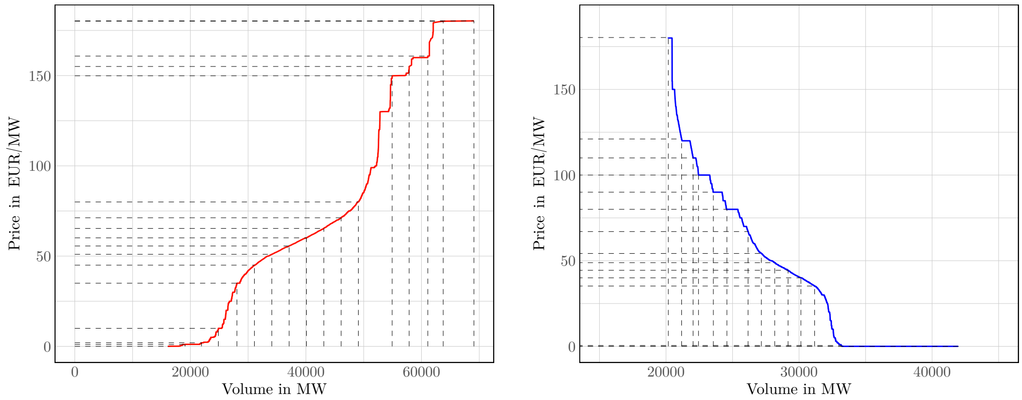
3.2.1 Price classes for bids

There are 18031 possible volumes on the full price grid P. And the usage of prices is due to burden computational. Therefore, the prices in P have to merge into a smaller amount of classes, and for the bids within the classes, it has to be assumed that the behave similarly over time.

When creating the price classes we consider the mean volume $\overline{V}_S(P)$ and $\overline{V}_D(P)$ as in equation (3), classes will be used to measure the importance of price at a later stage. Similarly to the definition of the price curve in equation (2), however, we are going to aggregate the mean supply and demand, therefore we obtain the mean supply and demand curve.

$$\overline{S}(P) = \sum_{p \in \mathbb{P}_S} \overline{V_S}(p) \text{for P} \in \mathbb{P}_{\mathbb{S}} \text{ and } \overline{D}(p) = \sum_{p \in \mathbb{P}_D} \overline{V_D}(p) \tag{4}$$

Figure 3: Mean Supply and Demand Curve $\overline{V_S}(P)$ and $\overline{V_D}(P)$ and class bounds using $V_S^* = 3000$ and $V_D^* = 1000$



(a) Mean Supply Curve

(b) Mean Demand Curve

For creating the price classes an amount of V_* is needed to set the bound for the average amount of volume that should represent each class, in the literature a V_* of 1000 has been used. However, we are going to use a $V_S^* = 3000$ and $V_D^* = 1000$, because the total amount is different between demand and supply. Given V_S^* and V_D^* we obtained a total of $M_S = 17$ and $M_D = 15$ classes for supply and Demand as shown in Figure 3, where the dashed line represent the price bound of each class. Price 0 and 180.30 have their class, whereas the other classes represent a group of bid price as follow, $\mathbb{P}_S(0) = 0$ and for example $\mathbb{P}_S(30) = 0.01, 0.02, ..., 29.99, 30$ and the associated volumes at time t to price classes $\mathbb{C}_S$ and $\mathbb{C}_D$.

$$X_{S,t}^{(c)} = \sum_{P \in \mathbb{P}_S(c)} V_{S,t}(P) \text{ for } c \in \mathbb{C}_S$$

$$X_{D,t}^{(c)} = \sum_{P \in \mathbb{P}_D(c)} V_{D,t}(P) forc \in \mathbb{C}_D$$

3.2.2 Time series model for bid classes

An slightly transformation of $X_{S,t}^{(c)}$ and $X_{D,t}^{(c)}$is include, we are going to consider day and hour as the time subscript as follow $X_{S,d,h}^{(c)}$ and $X_{D,d,h}^{(c)}$, in addition, we are also going to consider the the load forecast of solar, wind, and conventional, then we will have 5 additional process apart from the classes, $X_{price,t}$, $X_{volume,t}$, $X_{generation,t}$ $X_{solar,t}$ and $X_{wind,t}$, plus, similar to the transformation considered before for the classes, we also going to consider it for these five process $t = \{d, h\}$.

This model is also able to capture the weekday effect, nevertheless, for this, we must include a weekday dummy variable as follows:

$$W_k(d) = \begin{cases} 1, W(d) < k \\ 0, W(d) \geqslant k \end{cases}$$

Where $W(d)$ give the number corresponding to the day d, for Monday $k = 1$, and Sunday $k = 7$.

To fully represent the time series we should introduce the object $X_{d,h}$ as in the original paper.

$$\begin{aligned} X_{d,h} &= (X_{1,d,h}, ..., X_{M,d,h})' \\ &= ((X_{S,d,h}^{(c)})_{c \in \mathbb{C}_S}, (X_{D,d,h}^{(c)})_{c \in \mathbb{C}_D}, X_{price,d,h}, X_{volume,d,h}, X_{generation,d+1,h}, X_{solar,d+1,h}, X_{wind,d+1,h})' \end{aligned}$$

Generation, wind, and solar are going to be considered the planned class, they have the $d+1$ subscript because they are known one day in advance. the dimension of $X_{d,h}$ is $M = M_S + M_D + M_X$ ($M = 17 + 15 + 5 = 37$ given the used data), moreover, we subtract the mean in order to obtain a zero mean process $\mathbf{Y}_{d,h} = \mathbf{X}_{d,h} - \mu_h$ where, $\mu_h = \mathbb{E}(\mathbf{X}_{d,h})$, then the model will be as follows:

$$Y_{m,d,h} = \sum_{l=1}^{M} \sum_{l=1}^{24} \sum_{k \in I_{m,h}(l,j)} \phi_{m,h,l,j,k} Y_{l,d-k,j} + \sum_{k=2}^{7} \psi_{m,h,k} W_k(d) + \varepsilon_{m,d,h} \tag{5}$$

where $\phi_{m,h,l,j,k}$ represents the autoregressive part and $\psi_{m,h,k} W_k(d)$ represents the weekday effect, $I_{m,h}(l, j)$ the set of lags and $\varepsilon_{m,d,h}$ is the error term.

The set of lags $I_{m,h}(l,j)$ can be selected freely (Ziel and Steinert (2016)). In this Master Thesis, we have selected 30 lags (1 month) for the same hour in the target price class, for different hours we are only considering 7 lags (1 weak). For the other price class, we considered 7 lags for the same hour and only 1 lag for different hours, and finally, for the planned class we considered only 7 lags for the same hour. This can be represented as follows:

$$I_{m,h}(l,j) = \begin{cases} \{1,2,...,30\} & , m = l \text{ and } h = j \\ \{1,2,...,7\} & , (m = l \text{ and } h \neq j) \text{ or } (m \neq l \text{ and } h = j) \\ \{1\} & , m \neq l \text{ and } h \neq j \end{cases}$$

For the estimations of the parameter in (7) we are going to use the lasso estimation, which was introduced for the first time by (Tibshirani (1996)), is a penalized least squared and can be represented as follows:

$$\widehat{\widetilde{\beta}}_{m,h} = \arg \min_{\beta \in \mathbb{R}^{p_{m,h}}} \sum_{d=1}^{n} (\widetilde{Y}_{m,d,h} - \widehat{\mathbb{X}}_{m,d,h}\beta)^2 + \lambda_{m,h} \sum_{j=1}^{P_{m,h}} |\beta_j| \tag{6}$$

where $\lambda_{m,h} \geq 0$ is the penalty parameter. If $\lambda_{m,h} = 0$ it becomes a least square estimator and if $\lambda_{m,h}$ large enough we get $\widehat{\widetilde{\beta}}_{m,h} = 0$, $\lambda_{m,h}$ is selected by minimizing the Bayesian information criteria (BIC). Cross-validation techniques are also possible (Ziel and Steinert (2016)).

The estimated day-ahead point forecast can be represent as follows:

$$\hat{Y}_{m,n+1,h} = \sum_{l=1}^{M} \sum_{l=1}^{24} \sum_{k \in I_{m,h}(l,j)} \hat{\phi}_{m,h,l,j,k} Y_{l,n+1-k,j} + \sum_{k=2}^{7} \hat{\psi}_{m,h,k} W_k(n+1) \tag{7}$$

We obtained the bid volume forecast $\hat{X}_{1,n+1,h}, ..., \hat{X}_{M_S+M_D,n+1,h}$ and we add the subtracted the sample mean

3.2.3 Reconstructing bids and price curves

After we have compute the forecast $\hat{X}_{1,n+1,h}$ for each class $m \in \{1, ..., M_S + M_D\}$ and hour (h) but we have to model the proportion of the forecast bid volume $\hat{X}_{1,n+1,h}$ for each class. For example, if we forecast the sale volume of the price class ranging from 0 EUR/MWh to 30 EUR/MWh to be 1000 MW, we have to redistribute this volume over the different price levels within that class, e.g. 0.01, 0.02, ..., 30. However, not all the prices are bids at all. This step is important because for example, a 0.01 MW can have a huge impact on electricity prices. Then, the use of probabilities will be necessary.

Let $\pi_{S;d;h}(P)$ and $\pi_{D;d;h}(P)$ be the probabilities that $V_{S;d;h}(P)$ and $V_{D;d;h}(P)$ respectively is greater than zero, so there is actually a bid at this price. And we estimate them as the relative frequency in the sample. Furthermore, we assume proportionality within the bid prices in the classes for the mean volume $\overline{V}_S(P)$ and $\overline{V}_D(P)$ For reconstructed objects, we will use the accent ($\breve{}$), then, we can express the reconstructed volume $\breve{V}_{S,d,h}(P)$ and $\breve{V}_{D,d,h}(P)$ by

$$\breve{V}_{S,d,h} = \frac{R_S(P)\overline{V}_S(P)}{\sum_{Q\in\mathbb{P}_S}^{(c)} R_S(Q)\overline{V}_S} X_{S,d,h}^{(c)} \tag{8}$$

$$\breve{V}_{D,d,h} = \frac{R_D(P)\overline{V}_D(P)}{\underset{Q\in\mathbb{P}_D}{\overset{(c)}{}} R_D(Q)\overline{V}_D} X_{D,d,h}^{(c)} \tag{9}$$

However, instead of using Bernoulli random variables with probability $\pi_{S;d;h}(P)$ or $\pi_{D;d;h}(P)$ to see whether the prices will be considered. Instead, we are going to set a threshold of 1/12 (the price appears at least twice a month), then if $\hat{\pi}_{S;n+1;h}(P)$ or $\hat{\pi}_{D;n+1;h}(P)$ is higher than the threshold (1/12), R_S or R_D is going to be 1, otherwise zero.

After we have computed $\hat{\pi}_{S;n+1;h}(P)$ and $\hat{\pi}_{D;n+1;h}(P)$ we can estimate the supply and demand volumes $\breve{S}_{d,h}(P)$ and $\breve{D}_{d,h}(P)$ as follows:

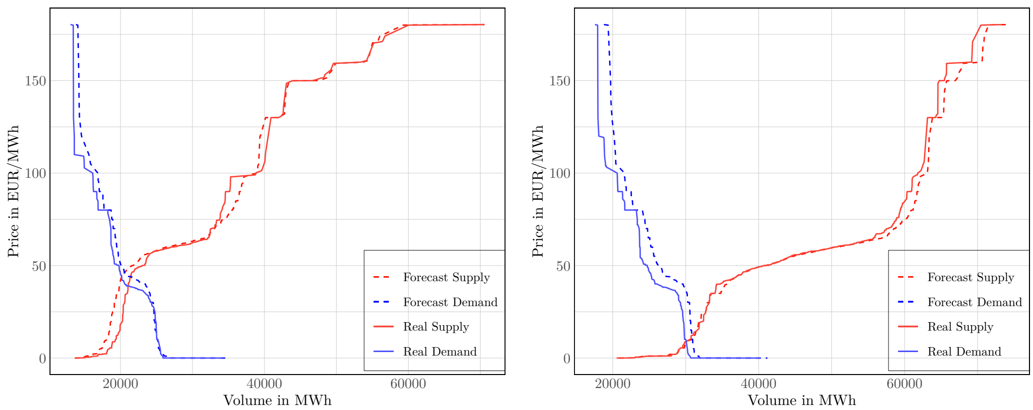
$$\breve{S}_{d,h}(P) = \sum_{p\in\breve{\mathbb{P}}_{S,d,h}} \breve{V}_{S,d,h}(p) \text{ for } P \in \breve{\mathbb{P}}_{S,d,h} \text{ and } \breve{D}_{d,h}(P) = \sum_{p\in\breve{\mathbb{P}}_{D,d,h}} \breve{V}_{D,d,h}(p) \text{ for } P \in \breve{\mathbb{P}}_{D,d,h} \tag{10}$$

where $\breve{\mathbb{P}}_{S,d,h}$ and $\breve{\mathbb{P}}_{D,d,h}$ represent the reconstructed bid prices and the intersection between both curves represent the clearing volume and prices.

4 Empirical results

To obtain the results we use a two-year rolling window. That is to say, if we want to forecast the date 01/01/2019, we have to use data from 01/01/2017 to 31/12/2018. In particular, for this Master Thesis, we forecast April 2019 and April 2020, in order to analyze the performance of the model by comparing 2019, where we do not observe any shock, and 2020, where we observe the Covid-19 pandemic and the associated lockdown policies, which were an important demand shock for the Spanish electricity market. We forecast the supply and demand curves for the whole month; however, for the sake of simplicity, in this report we only present figures related to the second Wednesday of April for both years. We use a standard day, trying to avoid week seasonality effects. Additionally, we compare one peak-hour **(20:00)** as well as a non peak-hour **(03:00)**, in order to observe the peak/off-peak effects.

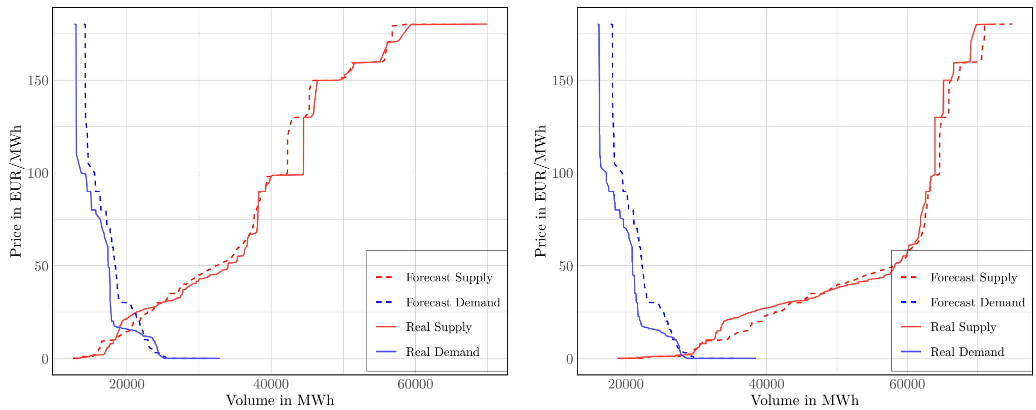
Figure 4: Forecast for April 17, 2019



(a) At 03:00 (b) At 20:00

In Figure 4 we plot the actual supply and demand curves (**solid red and blue lines, respectively**), as well as forecasted supply and demand curves (**dashed red and blue lines, respectively**) for an off-peak (**03:00**) hour and a peak hour(**20:00**). For both hours, the model seems to overestimate the actual demand curve. However, is hard to state visually at which hour the model behaves better. The supply forecast seems to have worse performance than the demand forecast during off-peak hours. First, it is sensitive to sharp price changes as can be observed in the price range. Second, something interesting happens that requires a further research: it seems that for prices below 60 EUR/MWh the model seems to underestimate the supply curve, whereas for prices above 60 EUR/MWh it somewhat overestimates. Lastly, the model seems to adjust better when the use of renewable energies increases, as it can be observed during the peak hours, where the horizontal segment is larger (i.e. more renewable energy).

Figure 5: Forecast for April 15, 2020



(a) At 03:00

(b) At 20:00

In Figure 5 we plot the actual supply and demand curves (**solid red and blue lines, respectively**), as well as the forecasted supply and demand curves (**dashed red and blue lines, respectively**) for an off-peak hour and peak hour, respectively. We observe that the model still overestimates demand, although we find that close to the marginal price it fits a little worse, yet its fit is pretty good. For the supply, the model does not fit as well as the demand and again, the model fits better when more renewable sources are used, as it can be seen during the peak hour, when the horizontal segment of the supply curve is larger (i.e. more renewable energy). Comparing the performance of the model in 2019 and 2020 we can state that the unexpected Covid-19 effect worsens the demand forecast.

So far we have seen that the model can represent the stepped behavior, which is characteristic of the energy market and it is quite satisfactory how well the curves fit, at least graphically. In any case, there is still room to improve and adjust the model to obtain more precise estimates. But to evaluate it we need to measure it quantitatively, so we will estimate some basic and standard measures of error.

To measure the performance of the model, measurements of common errors will be used, i.e. The absolute mean error (MAE_h) at time h and the root mean square error ($RMSE_h$) at time h, and instead of measuring the error with the price as Ziel and Steinert (2016) did in the original paper, we are going to measure the performance on forecasting the curves (Supply and

Demand) as can be seen in equation 11 and 12.

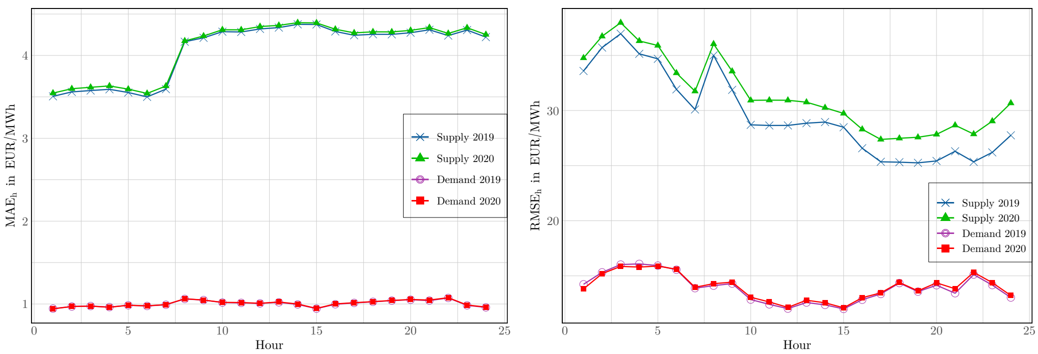
$$MAE_h = \frac{1}{(\#D)} \sum_{d \in D} |X_{volume,h} - \hat{X}_{volume,d,h}| \tag{11}$$

$$RMSE_h = \sqrt{\frac{1}{(\#D)} \sum_{d \in D} |X_{volume,h} - \hat{X}_{volume,d,h}|} \tag{12}$$

Both measures are suitable to compare point-forecast, however, in general, the MAE is more robust than the RMSE, as the latter is more sensitive to outliers (Ziel and Steinert (2016)).

In Figure 6 we plot the estimated **MAE** and **RMSE** for each hour and we observe interesting and at the same time concerning results. First, our MAE_h is concernedly low. Literature values are around 4:6, approximately; however, for the demand we obtain a value around 1. Furthermore, for the demand it is slightly constant, even under the demand shock observed in April 2020. On the other hand, the MAE_h of the supply is similar in magnitude to what we found in the literature, which is positive. However, something that needs further research may have happened at 07:00, causing a spike in the estimation and after that it behaves slightly constant. Finally, under the demand shock in 2020, the MAE_h is slightly higher.

Figure 6: MAE and RMSE in EUR/MWh of the X-Model for both forecast



(a) MAE_h (b) $RMSE_h$

When we estimate the $RMSE_h$ something similar happened. In the literature, we have found that this estimation is around 10, we obtain an $RMSE_h$ near this value, however, if we considered that both (MAE_h and $RMSE_h$) come from the same error and the second is ten times the first, and for the supply happened to be 7 more than 7 times. This needs to be revised, of course, there could be different reasons why we obtain such interesting errors. The first is the way we estimate the parameter R_s, instead of using a random Bernoulli probability, we used a threshold, and if the relative frequency was above that threshold, then, we considered that bid price, which could have caused a large number of outliers and affect the $RSME_h$ estimation. Another probability is that this model is not able to capture the ‘strategic” behavior of the generator, which is traduced in a higher error.

5 Conclusions

In this Master Thesis, we implement in the Spanish electricity market the X-Model proposed by Ziel and Steinert (2016), combining the insights of market structure models with econometric techniques. Instead of directly modeling the electricity price using past electricity prices, we model and utilize its true source: the sale and purchase curves of the electricity market. We incorporate not only several known features, such as seasonal behavior or renewable energy participation, but also other facts of the bidding structure, such as price clustering. The model is able to capture the non-linear behavior of the electricity price, which is especially useful for predicting huge price spikes.

In summary, we were able to forecast supply and demand curves for April 2019 and April 2020. The **X-Model** has shown promising results when forecasting both curves. In general, we observe a better fit for the demand than for the supply; however, this statement needs further research. On the other hand, the supply curve forecast is sensitive to the ”price spikes” which affect its fit. Additionally, the model in general -even on the demand shock- seems to have a better fit on the supply curve when the usage of the renewable source is larger. Furthermore, the model seems not to fully capture the “strategic” behavior of the generators, we also need further research in this regard. In addition, the estimated demand seems to always overestimate the curve, mean while for the supply curve a mixed effect is observed; however, is hard to say whether it overestimate or underestimate. Nevertheless, we can say that the supply seems to have a better fit when the usage of renewable source is larger, which is associated with the usage of wind and solar forecast.

When we estimate the performance of the model by estimating common error estimations (MAE and RMSE). A very low MAE was obtained for the demand and no significant difference is observed under the demand shock, which is surprising. For the supply, the magnitude of the estimation is in line with what we find in the literature. The RSME is the most concerning result of our Master Thesis and needs further research. Both estimations come from the same error, but the RSME is more sensitive to outliers than the MAE. We obtained an RSME 10 times higher than the MAE for the supply and on average 7.5 times for the supply, meaning that we could have many outliers. One of the hypotheses to explain this behaviour could be the way we reconstructed the volumes in equations 8 and 9. Instead of using a random Bernoulli distribution, we used a threshold of $\mathbf{1/12}$ to select the prices we are going to be considered as bids in the reconstructed curve. This approximation could lead to outliers and explain what we observe in the error measures. In this sense, on further research we plan to check our data and the differences of using both approximations.

Finally, this model has room for improvements. First, we will check the data and try to solve the problem of the outliers, and figure out how to manage them. Second, instead of doing a point forecast, it would be interesting to conduct a probabilistic forecast using the Bernoulli distribution. Additionally, another interesting approach is to distinguish between technologies (renewable, nuclear, hydro, etc), rather than focusing only in the aggregate supply curve. This is something the author of the original paper did not consider and that could have very interesting implications in terms of energy policy analysis. Third, we propose not only include the weakly effect but other seasonal effects. Fourth, we are going to use other benchmark models (naive models), in order to evaluate the gains of this model when forecasting supply and demand curves and also electricity prices. This is particularly important, in order to identify the advantage of using such a complex model for price forecasting instead of other simpler models. And finally, we would like to include additional regressors in the model, such as fuel costs or CO_2 allowance prices. By doing so, we could improve the fit of the curves and therefore obtain better results.

References

German Aneiros, Juan M Vilar, Ricardo Cao, and Antonio Munoz San Roque. Functional prediction for the residual demand in electricity spot markets. *IEEE Trans. Power Syst.*, 28 (4):4201–4208, 2013.

C Batlle and J Barquin. A strategic production costing model for electricity market price analysis. *IEEE Trans. Power Syst.*, 20(1):67–74, 2005.

René Carmona and Michael Coulon. A survey of commodity markets and structural models for electricity prices. In *Quantitative Energy Finance*, pages 41–83. Springer New York, New York, NY, 2014.

Aitor Ciarreta, Shahriyar Nasirov, and Carlos Silva. The development of market power in the spanish power generation sector: Perspectives after market liberalization. *Energy Policy*, 96: 700–710, 2016.

Michael Coulon, Christian Jacobsson, and Jonas Ströjby. Hourly resolution forward curves for power: Statistical modeling meets market fundamentals. In *Energy Pricing Models*, pages 147–193. Palgrave Macmillan US, New York, 2014.

Michael Eichler, Johan Sollie, and Dennis Tuerk. A new approach for modelling electricity spot prices based on supply and demand spreads. http://www.iot.ntnu.no/ef2012/backup/files/papers/53.pdf, 2012. Accessed: 2022-6-4.

Antonio García-Alcalde, Mariano Ventosa, Michel Rivier, Andrés Ramos, and Gregorio Relaño. Fitting electricity market models. a conjectural variations approach. https://pascua.iit.comillas.edu/aramos/papers/PSCCSeville_s12p03.pdf, 2002. Accessed: 2022-7-7.

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

Derk J Swider and Christoph Weber. Extended ARMA models for estimating price developments on day-ahead electricity markets. *Electric Power Syst. Res.*, 77(5-6):583–593, 2007.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc.*, 58(1): 267–288, 1996.

Mariano Ventosa, Álvaro Baíllo, Andrés Ramos, and Michel Rivier. Electricity market modeling trends. *Energy Policy*, 33(7):897–913, 2005.

Rafal Weron. *Modeling and forecasting electricity loads and prices: A statistical approach*. John Wiley & Sons, Nashville, TN, 2007.

Rafał Weron. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *Int. J. Forecast.*, 30(4):1030–1081, 2014.

Florian Ziel and Rick Steinert. Electricity price forecasting using sale and purchase curves: The X-Model. *Energy Econ.*, 59:435–454, 2016.