COMILLAS
UNIVERSIDAD PONTIFICIA
ICAI

MÁSTER EN INGENIERÍA EN TECNOLOGÍAS INDUSTRIALES E INDUSTRIA CONECTADA

TRABAJO FIN DE MÁSTER

Forecasting retail prices using a SARIMA model

Autor: Alicia Sanz De La Escalera

Director: José Pablo Chaves Ávila

José Pablo Chaves Ávila

Madrid

Contents

Tables

Figures

1. Abstract

The way consumers are having access to electricity and how retailers or large corporates are producing it is changing, and have been developing over time, mainly due to higher prices, political instabilities, and higher consumption levels. Within this scenario, multiple efforts have been implemented to create algorithms focused on generation, traders, and distributors to optimise their resources, maximising benefits to provide more cost-effective energy solutions.

This has led to an evolution of the electricity market, and the current situation, where the renewable energies have been increasing over the past few years considerably. In addition to that, electricity prices have also been increasing, mainly after the Covid/19 crisis, although now they are starting to get stablished. The factor of the constant change in electricity prices, generates uncertainty for both, retailers, and consumers, who have been incorporating long-term variables for the energy contracts, to reduce the volatility of the electricity market and reduce its impact on prices.

In this way, retailers are using long-term PPAs contracts, rather than acquiring all the electricity in the SPOT market to supply to its customers combining both sources of energy. The predictions done by retailers are focused on a 3-5 year horizon for the contract periods, to make models balanced between accuracy and longer predictions, as making predictions for 20 years from now, would make any model loosing accuracy [1].In fact, long-term contracts are profitable for retailersmainly when electricity prices within the SPOT market are high, however, if they reduce, long-term PPA contracts could be less cost-effective.

When estimating pricing for consumers, there can be considered two different ways, i) from one hand, by fixing a price per MWh, as in a long-term PPA contract, or ii) considering the volatility of the energy prices due to the market. To obtain a prediction of the volatility of the prices it is needed to take as reference the historical data and daily changes [2] [3].

Another key factor in the electricity sector, is the PVPC, which refers to Voluntary Price for Small Electricity consumer. Regulation that sets the methodology to establish voluntary prices for electricity consumers and the legal regime for contracting, configured as a dynamic price that internalises the volatility of the wholesale market price signal. [8]

The latest changes of the regulation of the PVPC are focused on ensuring transparency on the electricity prices, minimize the volatility of the prices for the small-end consumers, generate a competitive market, and promote the renewable energies, granting the access for all the users.

The latest regulatory measures implemented by PVPC have also incorporated long term component in the calculation of the electricity prices, together with the daily intraday market prices, in order to reduce the volatility and provide better stability of the prices.

In addition, using a SARIMA machine learning technique, future electricity prices are analysed for the following three months, showing how accurate is the model and its reliability, through the implementation of a programming code.

2. Introduction to the electricity market environment

Renewables energies have been increasing considerably during the past years, especially solar energy generation, which is having a high demand and its leading the energy transition, specially among households and small business, reducing the dependency of other energy sources. [4].

In the EU, solar PV energy generation increased from 25TWh in installed capacity in 2010, to 160TWh in 2021 [5]. Solar PV and onshore wind represented the biggest investment focus among all renewable generation sources globally. In addition, Western Europe remains one of the largest areas of investment for renewable energy. Self-consumption is in high demand because it allows lower energy prices, reducing monthly electricity instalments [5]

The investments in renewable energy have reached up to USD 280 billion over the last years, being 90% of the total investments done by the private sector. Companies are pushed to invest in renewables as they look for economic, social and environmental benefits, what translates into cost savings, long-term price stability and security of supply [6].

The European countries as shown in the Figure below, have been focusing and driving towards a more sustainable matrix, reducing the dependence on fuels, with ambitious objectives for the future, this also promotes technological innovation and the integration of new systems.

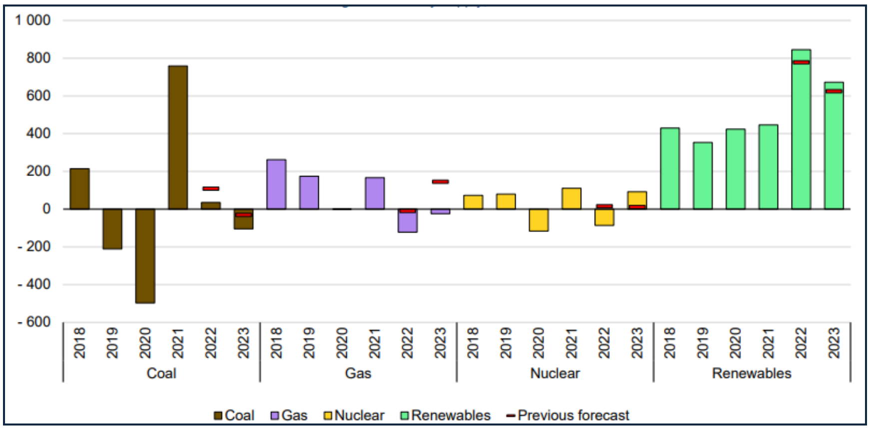

Figure 1: Evolution of energy sources.

Regarding coal, many countries across Europe have started to close their coal plants, while adding more taxes and more strict regulatory measures.

On the other hand, despite the gas can be cleaner compared to the coal, this source of energy still represents one of the sources that provides more CO_2 to the environment, in some countries that have been mainly focused on reducing the coal, the consumption of gas has increased slightly.

Nuclear energy provides a stable and CO_2 emission-free source but faces public opposition and waste management challenges. Countries as France, remain and keep a high nuclear dependency, however, others as Germany, is in the process of removing completely this source gradually.

From 2020 onwards, renewable energies have been accounting for almost 40% of the total energies across Europe, with the main sources of generation being biomass, geothermal and marine energy. [7]

3. PVPC regulatory changes

PVPC has recently introduced changes to reduce the volatility of the electricity prices for small consumers, by implementing mainly, among others, the following:

1. Introduction of a mixed Price: from January 2023, the PVPC has included a long-term forward component, that goes together with the daily hourly price of the SPOT market. This change aims to reduce the volatility of the PVPC and offer a better pricing stability.
2. Long-term components: apart from the SPOT market, long-term variables are introduced, estimated as a weighted average of 25%, 40% and 100%, during the first, second and third year respectively.

In addition, PVPC has incorporated a new structure of tolls and charges, by providing a classification of time periods for every access tariff (peak, flat and valley). All consumers with less than 15KW of contracted power will have hourly discrimination in three different periods. [8]

PVPC recent changes also promotes the renewable energy, and its integration within the energetic mix, looking for reducing the dependence of fuels and gas, and its associated volatility costs. For this, PVPC implemented measures to promote the costs-saving in electricity.

All the measures considered and implemented, are with the aim of:

1. Reducing volatility prices for the small consumer, due to the highly volatility of the daily market.
2. Adaptation to new European regulatory changes.

3. To promote sustainability and to grant transparency and competitiveness.

4. Statistical analysis: Machine Learning techniques

Machine learning techniques are widely used in statistical analysis to uncover patterns, make predictions, and infer relationships within data. The most common techniques applied to predict data, specifically designed to analyse time-ordered data points, would be SARIMA, Exponential Smoothing, LSTM networks, or others such as Transfer function time series or non-linear regression models, using multi-linear perceptron models.

In particular, the transfer function is a system or mathematical function that represents the output based on its corresponding inputs adjusting data that varies over time. To predict and fine-tune the model accurately, it is essential to make the series stationary.

Multi-linear perceptron models, which could also be used for classification purposes, is based on a neutral network with one hidden layer. In this case, the inputs are given by explanatory variables chosen, and there will only a reduced number of layers, as the outcome. [9]

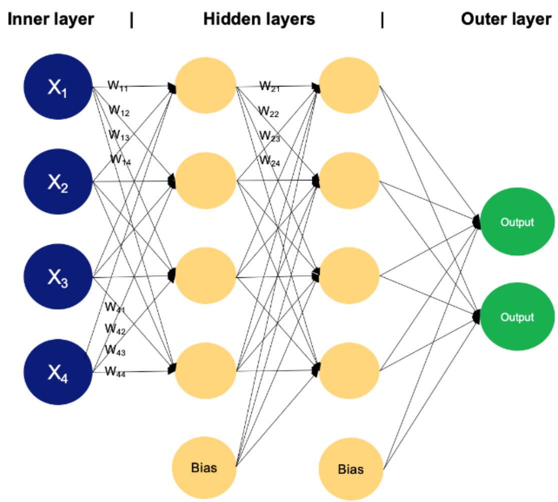

Figure 2: Multi-linear perceptron models

Finally, SARIMA model has been used to analyse and predict the expected values for the electricity prices. SARIMA is the machine learning technique mainly used when the data has a seasonal behaviour and component. SARIMA, refers to Seasonal Autoregressive Integrated Moving Average, which is an extension of ARIMA, commonly used in statistics and econometrics, fitted to time series data, to predict future data points. The SARIMA function considers three hyperparameters to define the autoregression, differencing and moving average, relevant for series of data where there is seasonality. [10]

SARIMA has the following components, both seasonal and non-seasonal:
From one side, the ARIMA components:

- AR (Auto Regressive): This component uses the dependency between an observation and a number of lagged observations (previous values).
- I (Integrated): This part involves differencing the raw observations to make the time series stationary (i.e., to remove trends and seasonality).
- MA (Moving Average): This component uses the dependency between an observation and a residual error from a moving average model applied to lagged observations.

From the other, the seasonal components:

- Seasonal AR (P): Seasonal autoregressive order.
- Seasonal Differences (D): Seasonal differencing order to make the series stationary.
- Seasonal MA (Q): Seasonal moving average order.
- Seasonal Period (S): The number of periods in a season (e.g., 12 for monthly data with yearly seasonality)

The model determines if the time series is stationary. If not, apply differencing to make it stationary. Identify the order of seasonal differencing required. It later uses statistical techniques to estimate the parameters of the AR, MA, seasonal AR, and seasonal MA components, to later check the residuals of the model to ensure they resemble white noise (i.e., no pattern, indicating a good fit), and finally produce the forecast.

These techniques can be used individually or in combination to analyse data, uncover insights, and make predictions. Each technique has its strengths and weaknesses, and the choice of technique depends on the nature of the data and the specific problem being addressed.

5. Electricity prices prediction based on machine learning techniques

Considering machine learning techniques, the electricity prices in Spain are forecasted considering as input data the daily price load of the wind and its demand.

As first step, a set of data has been trained, also to analyse the behaviour of the information, and analyse wether it is stationary or seasonally, where stationary variables refer to data with constant variance, and constant correlation over time, while seasonally variables refer to data with periodic fluctuations, and non-constant variance.

As the inputs are electricity related, showing seasonality, changing over time, the average of the data set and the variance are not constant in time, what results in a higher probability to differentiate the average, and what increases the need of running more than one scenario to compare the analysis.

5.2. SARIMA model

SARIMA model has been used to forecast the electricity prices. The model considers also a combination of ARIMA, finding the optimized model parameters based on the best fit between the predictions and the test data.

The evaluation metrics considered used to fit and select the best model parameters, in order to create an accurate analysis, are ME, which stands for Mean Error; MAE, Mean Square Error; MAPE, Mean Absolute Percent Error; MAE, Mean Absolute Error; MPE, Mean Percent Error; RMSE which stands for Root Mean Square Error, and finally ACF that refers to Auto- Correlation function. [11]

As shown in the table below, the parameters of the training error measure, are as follows:

SARIMA	ME	RMSE	MAE	MPE	MAPE	ACF1
Training	0.0992	4.278	3.1590	-1.1350	9.5440	0.0057

Table 1: Training error measures

The SARIMA model meets the target values, as the noise is considered white noise, with a reasonable low value for RMSE. Considering these factors, it can be assumed that the model is reliable.

5.3. Output result

After analysing the data, it has been observed that the standard deviation does not depend on the level of series, and so it is not needed to apply any Box-Cox transformation of the initial set of data As it can be seen in the figure below, it is observed a partial correlation between the variables. Box-cox transformation is mainly applied to stabilize the variance and make the data more closely to a normal distribution, normally done to improve the performance and accuracy of the prediction models.

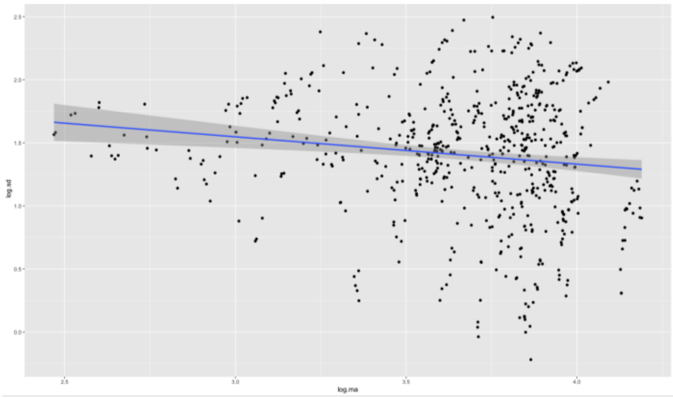

Figure 3: Partial correlation between variables

In the process of analysing the data, and adjusting the parameters of the model, all the coefficients that are being used are significant and the residual model does not exist as a component, but corresponds as white noise, what makes that the prediction model is highly reliable. As it can be seen in the figures below, both the data provided and the trained data merge being in a good agreement, showing that the parameters have been adjusted correctly.

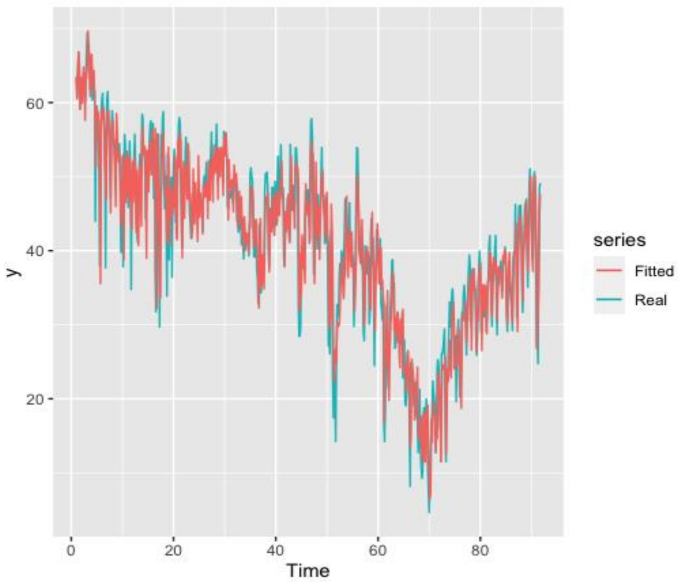

Figure 4: Fitted data vs real data used.

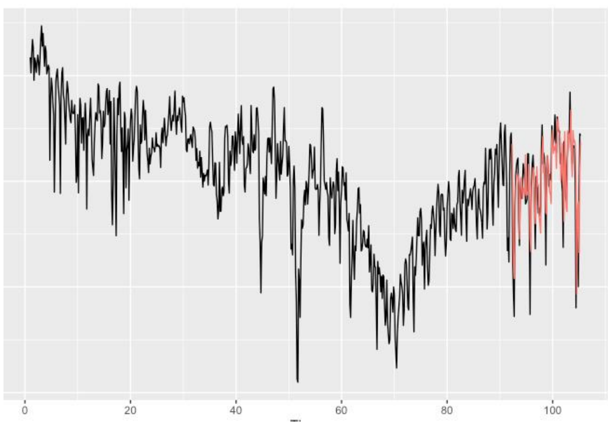

Figure 5: Predicted values for the following months.

Finally, the following measures have been obtained for the tests set.

SARIMA	ME	RMSE	MAE	MPE	MAPE
Test	0.5715	5.0484	4.7840	0.9844	9.7580

Table 2: Test error measures

The differences are mainly due to the fact that there is a high variance, and due to the differences between the data of the last three months compared to the rest of the data set. In addition, between tests and training data, it has also to be considered any external influence that may vary the demand, and so the inputs considerably.

6. Conclusions

In summary, the discussion has centred on the evolving electricity market in Europe, highlighting a new reform designed to accelerate the consumption of renewable energy. This reform aims to reduce reliance on gas, thereby shielding consumers from the volatility of fuel prices and potential market manipulation, while also making the industry more competitive and environmentally friendly. The analysis focused on achieving energy efficiency for end consumers from a retailer's perspective, who must optimize costs and grow their market share.

In addition, it has been checked that between all the machine learning techniques that can be applied to estimate data, SARIMA model can help predicting future values, in a very reliable way and accurately, if parameters are adjusted correctly, and so it can be very helpful to predict future electricity values. The test error values obtained could be considered acceptable, as the RMSE shown is relatively low, and remain close to the training values.

Bibliography

[1] H. Edwardes, “Long-term PPA deals being squeezed by price volatility,” S&P Global Ratings., February 2022.

[2] C. Agricole, “ Project Bond focus,” January 2022.

[3] KBRA, “Solar ABS. Luminace ABS,” 2022.

[4] H. Edwaardes-Evans, “European PPA offers relatively stable despite rising power prices,” S&P Global Ratings, 2022.

[5] KBRA, “ESG, Project Finance and Infrastructure,” 2022.

[6] U. Nations, “Global Sustainable Development Report,” 2022.

[7] Eurostat, Shedding light on energy in the EU, 2023.

[8] BOE, “Electricity prices methodology for the small consumer 216/2014,” Boletin Oficial del Estado, 2014-2022.

[9] IEEE, Application of multi-linear regression models and machine learning techniques for online voltage stability margin estimation, 2020.

[10] Mathworks, Machine Learning, how it works, why it matters, 2022.

[11] Forecasting the dynamics of cumulative COVID-19 cases (confirmed, recovered and deaths) for top-16 countries using statistical machine learning models: Auto-Regressive Integrated Moving Average (ARIMA) and Seasonal Auto-Regressive Integrated Moving Average, 2021.

[12] Mathworks, Machine Learning, how it works, why it matters, 2024.