ISSN 2063-5346

# ENHANCED ARIMA APPROACH OF ELECTRICITY PRICE FORECASTING

**Ali Abbasov**

**Article History: Received: 01.02.2023 Revised: 07.03.2023 Accepted: 10.04.2023**

**Abstract**

Forecasting the price of electricity has grown into an increasingly crucial part of the day-to-day operations of power providers. The efficacy of energy market participants, including producers as well as purchasers, may be increased by using accurate forecasting models. The process of investment planning also involves the consideration of price in a significant way. Using the well-known ARIMA approach, which is used for analysing and predicting time series data. The model is implemented using time series that are comprised of the day-ahead cost of electricity obtained from the EPEX energy exchanges. Forecasting the price of electrical power is an important responsibility in the energy business because it enables market players to make educated choices with regard to the trading of energy, the management of risks, and the distribution of resources. In the last few years, machine learning and additional statistical approaches have seen widespread use in an effort to enhance the accuracy of power price forecasts. This article presents and examines the effectiveness of a variety of machine learning and mathematical approaches. Additionally, the study discusses the implications of these findings. We give advice for picking the technique that is most appropriate for certain forecasting jobs after analysing the benefits and drawbacks of a variety of methodologies and comparing them. According to the findings, classical statistical models are outperformed in terms of accuracy by models based on machine learning that include artificial neural networks, support vector models, and random forests. The amount of volatility, seasonality, and load patterns are all factors that should be considered when choosing a model; nevertheless, the decision ultimately rests with the unique features of the electrical market. In conclusion, we analyse the shortcomings of the currently available methods and provide some suggestions for new lines of inquiry that may be pursued in the future to improve the precision and dependability of forecasts of future power prices

Lecturer, Niagara College Toronto business school, Canada

**DOI:10.31838/ecb/2023.12.s1-B.344**

## I. Introduction

The price of electricity consistently ranks unpredictable of all merchandises. The daily regular fluctuation in the spot value of electricity may be up to fifty percent, but the typical day change for other staples can be up to five percent. Numerous market participants, including generators, traders, suppliers, and end users (especially big industrial companies), are dependent on fluctuations in the price of electricity. These market participants include major industrial clients in particular. It is abundantly clear that it is quite crucial for them to possess precise forecasting models for the pricing of power.

Utilising information obtained through the European Electric Exchange as the study's standard power market will allow the researchers to accomplish their main goal of forecasting the day-ahead price for electricity. EEX collaborates with Power Next SA, which is its equivalent in France. EEX is a stakeholder has its headquarters in Paris and handles the infamous Spot Markets for short-term transactions in electrical to the benefit of Germany, France, Austria, and Switzerland. EEX is also a shareholder through the joint enterprise EPEX Spot SE [1]. EEX owns 50% of the company's shares. A day-ahead market is what is known as a spot market in the electricity industry. The distribution of physical energy is often covered by an average hourly contract when dealing with spot contracts. A secret bidding process that takes place once every 24 hours serves as the basis for making the decision.

The purpose of this experiment is to test the hypothesis that ARIMA models are capable of accurately forecasting day-ahead power prices. The incorporation of sources of renewable power as well as the development of new technologies are driving a huge transition in the energy business, which is now in the midst of such a revolution. An accurate estimate of future power costs is one of the most significant obstacles that market players in the energy industry must overcome. Predicting the price of electricity is essential for the efficient operation of electrical markets, which includes managing hazards, the allocation of resources, and trading techniques. Anticipating the price of energy has been done using a wide variety of classic statistical as well as economic modelling techniques over the course of many years. On the other hand, these models are limited in their ability to capture the many non-linear linkages and patterns that exist in power markets.

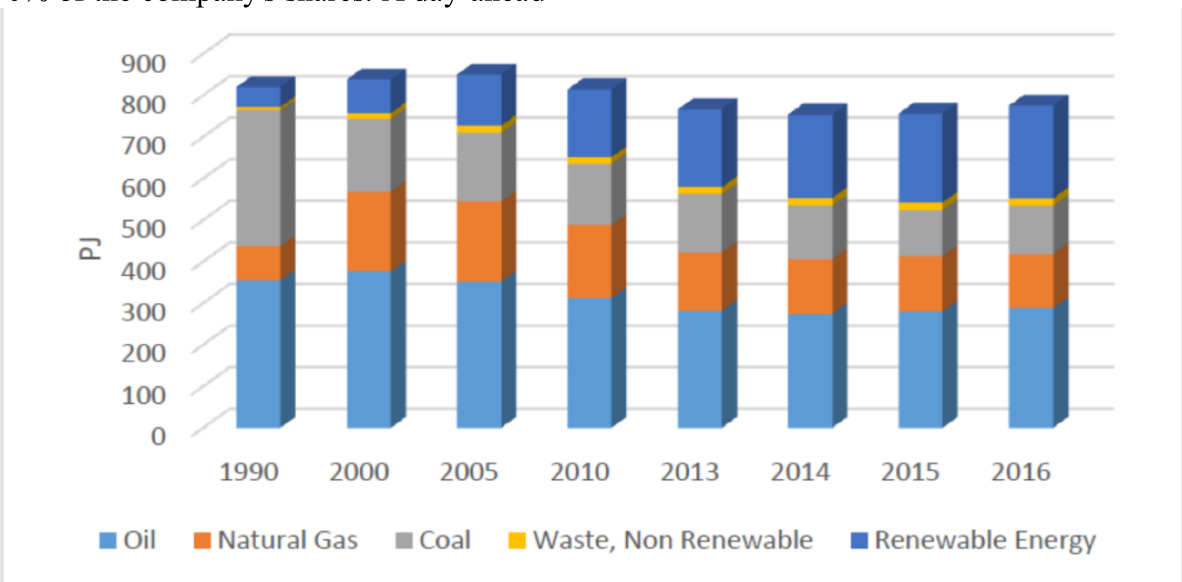



Fig.1. Energy consumption

As a result of its capacity to recognise intricate links and patterns within the data, the use of machine learning methods in the process of estimating future power prices has, as of late, been receiving an increasing amount of attention. For the purpose of power price forecasting, ARIMA models in particular, along with other Learning algorithms, supported, as well as random forests are examples of machine learning techniques, have seen widespread use. These mathematical models have shown encouraging results in terms of their accuracy, speed, and flexibility in a variety of different situations involving the power market. The purpose of this study is to give a complete assessment of the literature on the subject of energy price predictions using ARIMA models as well as machine learning methods. We examine the benefits and drawbacks of different models, evaluate how well they work, and then offer suggestions on how to choose the approach that is the best fit for certain forecasting endeavours. We also emphasise the benefits and problems associated with using machine learning approaches to predict power prices, as well as recommend future studies on topics based on these findings. In general, this study adds to a better knowledge of the state-of-the-art approaches for power price forecasting as well as its potential for boosting the sustainability and efficiency of electrical power markets. This awareness is one of the main goals of the research presented here. ARIMA models have actually been used in the past for the purpose of pricing forecasting; however, these models are typically straightforward and rely on a limited number of observations, often ranging from three weeks to an entire year. In this piece of research, the initial dataset has a total of 3836 observations over 10 years. An experienced modeller is used in order to locate the ARIMA model that best fits the data [2].

## II. Literature Review

Electricity price forecasting is an important area of research for power system planning and operation. Accurate forecasting of electricity prices can help utilities and market participants make informed decisions on generation, transmission, and trading strategies. In this literature review, we will discuss some of the key research papers and approaches to electricity price forecasting.

1. Time-series forecasting approaches: Models based on time-series forecasting have been used often in the process of estimating future prices of electricity. In the field of short-term forecasting, autoregressive integrated moving average (ARIMA) techniques are often used. These models consider historical prices and identify patterns in the time-series data. Some recent studies have also used machine learning algorithms such as support vector regression (SVR) and artificial neural networks (ANN) for forecasting.
2. Weather-based forecasting: Weather-based forecasting models take into account the impact of weather on electricity demand and supply. These models use historical weather data along with electricity demand and supply data to predict future prices. Some studies have also used ensemble forecasting techniques to improve the accuracy of weather-based models.
3. Market-based forecasting: Market-based forecasting models use information on the behaviour of market participants such as supply and demand, market rules, and regulations. These models are generally used for long-term forecasting and consider factors such as fuel prices, investment costs, and environmental policies. Some studies have also used game

theory-based approaches to model the behaviour of market participants.

4. Hybrid forecasting approaches: Hybrid forecasting approaches combine two or more forecasting methods to improve the accuracy of the forecasts. For example, some studies have used a combination of ARIMA and ANN models for short-term forecasting. Other studies have used a combination of market-based and time-series forecasting models for long-term forecasting.

## III. Methodology

The methodology for electricity price forecasting can vary depending on the timeframe and the specific requirements of the application. However, in general, the following steps are typically followed:

1. Data collection: The first step in electricity price forecasting is to collect relevant data. This includes historical price data as well as data on weather, demand, supply, and other relevant factors. The data should be reliable, complete, and of high quality.
2. Data pre-processing: Once the data is collected, it needs to be pre-processed to remove any noise or outliers and to transform the data into a suitable format for modelling. This may involve data cleaning, normalisation, and scaling.
3. Feature selection: Next, the relevant features for forecasting need to be identified. This may involve feature engineering to create new features or selecting relevant features from a larger set.
4. Model selection: Once the features are identified, a suitable forecasting model needs to be selected.
5. Forecasting: Finally, the model can be used to make forecasts for future electricity prices. The forecasts can be updated as new data becomes available, and the model can be retrained periodically to improve its performance.

**Cross Industry Standard Process for Data Mining**

CRISP-DM is an established protocol that is widely used to represent the life phases of a method for data mining. As can be seen in Figure 1, the life cycle is broken up into six distinct stages.

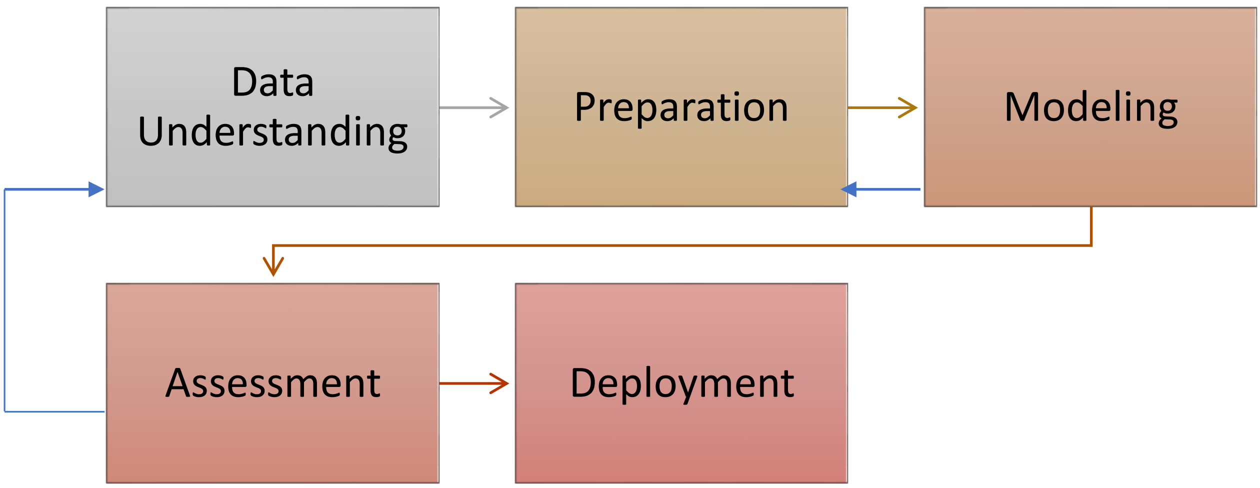



Fig. 2. CRISP-DM benchmark model's rounds of development

The first phase of the process is called "Business understanding," as well as consists of translating the expertise of the company into a data mining issue statement. Understanding the information, examining the data sets, and receiving preliminary insight into the information in order to formulate hypotheses comprises the second step. During the third phase of the process, which is called "Data preparation," we not only produce a final dataset but also execute any required data cleaning along with transformations. In the fourth step, known as "Modelling," we put several modelling strategies to use in an effort to address the data mining challenge. Evaluation is the name of the fifth stage, which is also a very significant step. During this step, we check to see how accurate the model is, as well as whether necessary, we may make any necessary adjustments to the simulation before we use it. The next phase is called "Deployment," and it involves applying the model to actual data [3].

The outcome of each stage will inform our choice on whether to go ahead or backwards after that stage. Iterations are the building blocks of this process.

In the fields of data mining as well as predictive modelling, one approach that sees widespread use is known as the Cross- Industries Standard Process for Data Mining, or CRISP-DM. It offers a methodical strategy for the completion of data mining projects and is divided into the following six stages: company understanding, information comprehension, preparing the information, modelling, assessment, and deployment. The ARIMA models as well as the machine learning algorithms may both be used in the prediction of future power prices with the help of this technique. In the following paragraphs, we will discuss how the CRISP-DM methodology may be used to provide accurate projections for future power prices.

- Business Understanding: During this stage of the project, you will be tasked with outlining its goals and specifications. The goal of predicting the price of electricity is to properly estimate its future price, which may then be utilised for risk management, resource allocation, and trading strategy. In order to construct accurate forecasting models, it is vital to have a thorough understanding of the features of the electrical market, such as its seasonality, volatility, and load patterns.
- Data Understanding: During this phase, the data sources are tracked down and examined so that a deeper comprehension of the data may be achieved. It is crucial to have past data on power prices when attempting to estimate future prices. It is important to do an analysis on this data in order to spot any patterns, seasonality, outliers, or missing numbers. It is also essential to have a solid grasp of the data distribution as well as the correlations that exist between the various variables.
- Data Preparation: Data preparation for predicting energy prices requires eliminating outliers as well as missing values, converting the data to ensure stationarity, and constructing additional attributes such as lagged variables as well as shifting averages. Other features may also be created. Modelling:
- During this stage, a variety of models are constructed and tested using the information that has been gathered. When dealing with ARIMA models, the model to use is decided upon by using the AIC as well as BIC criteria, as well as the limits of the perfect are estimated by applying the greatest probability estimation. Cross-validation methods are used in

order to train and assess a number of different which are used for machine learning models.

- Evaluation: In this step, the performance of the models that have been constructed. Comparisons of the models should be made based on how well they perform on the test data, and the model that proves to be the most successful should be chosen for deployment.
- Deployment: During this phase, the model that was chosen is put into action and incorporated into the infrastructure of the power market. To maintain its precision and dependability, the model has to have its data frequently refreshed with new information.

The CRISP-DM approach may be used in the context of predicting energy prices, and it can do so with either ARIMA models or machine learning algorithms. The methodology offers a systematic strategy for generating trustworthy and accurate forecasting systems, which may seem useful for the successful management of energy markets. This can be accomplished by employing the methodology. Nevertheless, peculiarities of the electrical market as well as to use the most relevant modelling methodologies.

### A. Business and data understanding

In this analysis. In order to accurately predict future power prices using ARIMA models and machine learning algorithms, one must first have a comprehensive knowledge of the electrical market and the data that is readily accessible for examination. In this part of the article, we will talk about the business and data interpretation stages of the CRISP-DM approach for predicting future power prices.

Business Understanding: During the business understanding phase of the CRISP-DM process, you will be tasked with identifying the project's goals and prerequisites. Accurately predicting future prices of electricity is the goal of the process known as "forecasting," which is done in the framework of the industry of "electricity price forecasting." Participants in the energy market use this information in order to manage risks, allocate resources, and make trade choices.

To be successful in reaching this goal, it is essential to have in-depth electrical market. The market for electricity is a convoluted system that encompasses all stages of the power supply chain: production, transmission, and distribution to consumers. The dynamic in the market is what ultimately decides the price of power. There are a number of factors that may have a substantial influence on the supply and demand for energy and, as a result, the price of power. Some of these factors include weather conditions, fuel costs, and regulatory laws.

In addition to having a solid grasp of the electrical industry, it is of the utmost importance to recognise the many stakeholders and the expectations they have. Energy producers, retailers, traders, and regulators are all potential parties to consider when attempting to anticipate prices for power. These many stakeholders have varying expectations and goals, and the forecasting model needs to be crafted in such a way as to cater to each of their individual requirements.
Data Understanding: During the data understanding phase of the CRISP-DM approach, you will be tasked with locating and examining all of the data that is accessible for analysis. It is crucial to have past data on power prices when attempting to estimate future prices. These statistics are available from a wide variety of sources, including the operators of the power market, the energy exchanges, and governmental organisations.

It is recommended to do an analysis of the data get features of the power market and

the quality of the data. During the analysis, you should look for patterns and trends as well as seasonality, outliers, and missing numbers. In order to construct reliable forecasting models data distribution correlations that exist between the various variables.
Exogenous variables as well as endogenous variables are the two categories that may be used to classify the data that pertains to the energy market. The term "exogenous variables" refers to any external factors, such as weather conditions, fuel costs, or regulatory laws, that have the potential to influence the market for electricity. Endogenous components are internal elements that are peculiar to a market. For example, the demand for power, the supply of electricity, and the transmission capacity are all examples of endogenous variables.

During the process of interpreting the data, you will also need to identify any problems with the data's quality and devise techniques that are suitable for fixing them. Examples of methods that may be used to impute missing data include interpolation and regression analysis. Either the data may be altered to lessen the influence of the outliers on the model, or the outliers themselves can be discovered and deleted.

When it comes to constructing accurate and dependable models for predicting the price of energy using ARIMA as well as machine learning algorithms, the business as well as information understanding stages of the CRISP-DM approach are very necessary. For the purpose of constructing models that satisfy the expectations of the stakeholders as well as provide significant insights into the industry of electricity, it is vital to have a thorough knowledge of the electricity market, the partners involved in the electricity market, and the data that is readily accessible for analysis.

The data from the trading intervals is used to generate a time series containing daily arithmetic means, which in turn produces 3836 occurrences for the aforementioned reference market [4].

### B. Data preparation

The present time series, like many others measuring macroeconomic variables, is either integrated or nonstationary. In order to get information ready for statistical modelling, series are converted to stationarity. This may be done by determining the natural logarithm of the series, taking the difference between the series and each other, or collecting the residuals of the data through a regression 4. The phase of the project titled "Electricity Price Forecasting," known as "Data Processing," is an essential part of the project since it is during this phase that the information and data are cleaned, pre-processed, and converted in order to get them ready for analysis. In this part of the article, we will talk about the data processing phase of the CRISP-DM approach for predicting the price of energy using ARIMA models as well as machine learning algorithms.

Data Cleaning: The act referred to as "data cleaning." In the context of estimating future prices for energy, "data cleaning" entails eliminating any outliers, supplying missing figures, and rectifying any mistakes that may be present in the information. Because the quality of the data may have a considerable influence on the correctness of the model used for prediction, it is very important to guarantee that the information is correct and has not been tampered with in any way.

Data Pre-Processing: The transformation of the data into a format that is better suited for analysis is what is meant by "data pre-processing." Techniques like normalisation, scaling, and feature engineering are involved in this process. In order to increase the efficiency of the algorithms used for machine learning, normalisation entails rescaling the information to fit within a predetermined range, such as between 0 and 1. The data are normalised by the process of scaling by first removing the mean, followed by dividing by the standard deviation, or SD. The process of choosing or developing additional factors that are related to the issue at hand and have the potential to enhance the effectiveness of the forecasting system is known as "feature engineering."

Feature Selection: The technique of determining which variables are most essential to include in the model for prediction is referred to as "feature selection." In the context of predicting the price of energy, this entails picking the factors that are most closely connected with the cost of electricity, such as the current state of the environment, the pricing of fuel, and the amount of demand for power. The complexity of the data must be reduced, and unimportant variables must be eliminated, before the quality of the forecasting model can be improved. This may be accomplished by feature selection.

Time-Series Analysis: When it comes to predicting the price of energy using ARIMA models, one of the most important techniques is time-series analysis. The breakdown of a time sequence into its trend, seasonality, and residual components is a necessary step throughout the time-series analysis process. The component known as the trend describes the behaviour of a given series over a longer period of time, while the seasonal component is responsible for identifying recurrent patterns in the information itself. The noise or random variations in the information that are unresolved by patterns or seasonally make up what is known as the remainder component of the analysis. In order to create effective forecasting models utilising ARIMA, it is vital to have an adequate grasp of the time-series components. The procedure of information splitting involves dividing the information into sets the fact that can be utilised later for school, certifying, as well as testing purposes respectively. The model that predicts outcomes is calibrated with the help of the initial collection of information, while the data from validation can be utilised to make adjustments to improve the system hyperparameters and pick the model that performs best; and the test data set is deployed to assess the outcome of the final model. All of these sets are used in sequence. It is vital to partition the data in order to avoid overfitting the predictive model and to make certain that the model is able to generalise adequately to fresh data.

The part of the project known as "information preparation" is an extremely important stage since it is at this phase that the data are cleansed, pre-processed, and altered in order to get them ready for analysis. Finding and fixing mistakes, contradictions, and absences of values in the information is what's referred to as "data cleaning," while "data pre-processing" refers to the process of converting the data so that it's more suited for analysis, which is known as "data transformation." Developing reliable projection models using ARIMA requires time-series analysis as a foundational component. Feature selection refers to the process of determining which variables are most helpful for incorporating into the forecasting model. In the last step, data splitting, the data is separated into test, validation, and training sets. This is done to prevent the model from becoming too specific to the existing data and to guarantee that it generalises well to freshly collected information.

After preparing the information for the software tool SPSS, we conduct an analysis of the primary features of the time series, which include the trend, cycle, and season. We discover that the chronology in question is non-stationary, which leads us to conclude that the use of a logarithmic transformation is necessary [5].

### C. Modelling – Box and Jenkins model

ARIMA is a technique that was developed according to Jenkins and Box 5 and that we utilise for modelling reasons. The ones that follow are a formulation of the ARIMA approach in its general form:

$$(B)p_t= \quad (B) \quad (1)$$

wherein (B) as well as (B) are variables of the backshift operators B:B lp = pt-1, and t is the magnitude of the error phrase. The price at time t is denoted by pt. The ARIMA model categories are presented using the conventional notation of ARIMA (p,d,q), as well as their seasonally equivalents, denoted by the letters P, D, and Q, respectively.

Autoregressive (p). The total number of consecutive orders that are autoregressive models in the regression model The autonomous order of a series specifies which earlier values in the sequence can be employed to make forecasts about future values [6].

Difference (d). Identifies the sequence in which differencing operations are carried out on the series prior to the application of estimation models. Differentiation is required when there is evidence of trends (data series that exhibit trends are often nonstationary, while ARIMA modelling assumes that the data are constant), and the method is used to nullify the impact of these trends. The degree of a series' trend is reflected in the sequence in which differencing is performed: first-order differentiation accommodates linear developments, second-order distinction compensates for polynomial trends, and so on. moving average (q). The total number of iterations of the moving average used in the calculation of the model Rules for moving averages indicate how departures from the collection's mean for earlier values are utilised to anticipate the values of the series at the present time.

Within the software programme Python is a feature known as the Expert Modeller, which may be used to do an automated search for the model that provides the best fit for every associated series. If multiple independent variables (predictors) are supplied, the expert modeller chooses only those variables that demonstrate a statistically significant correlation with the series being modelled (the dependent series). These variables are then included in the ARIMA model. We also include the automated identification of outliers in our specifications. Expert modellers evaluate seasonal models. This option will not become active until the frequency has been specified for the dataset that is currently being used. The expert modeller additionally incorporates an ongoing model.

The Box and Jenkins model, which also goes by a well-known method for analysing time series that is used in the process of predicting the price of energy. The autoregression (AR) model, the integration (I) model, and the moving average (MA) model are the three elements that make up the ARIMA model. Utilising the ARIMA model, we will explore the modelling portion of the CRISP-DM approach for estimating future prices of energy in this part.

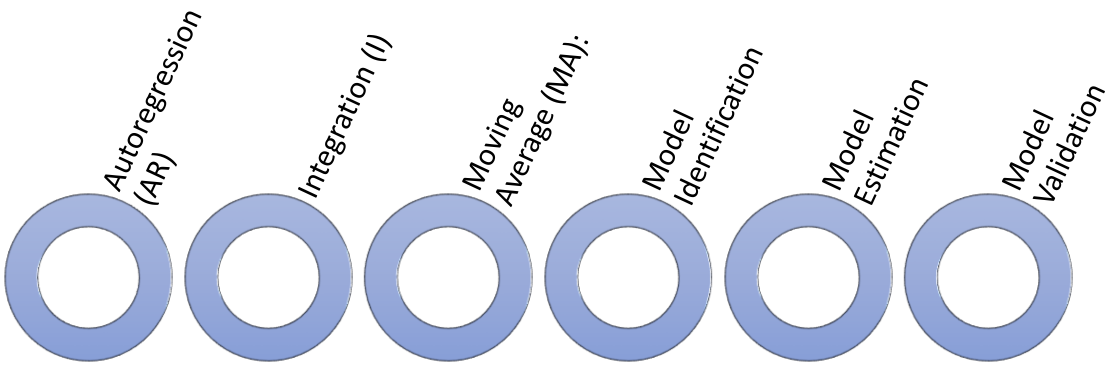



Fig.3. Various steps for evaluation

- Autoregression (AR): The present value of a parameter undergoes regression on its historical values, and this is the process that is referred to as autoregression. When it comes to projecting the price of energy, this entails doing a regression of the present price on its historical values. The number of historical values that should be included in the equation used for regression is decided by the order in which they appear in the autoregressive model component, which is indicated by the symbol p. The trend portion of the duration series is captured by the autoregressive model aspect of the analysis.
- Integration (I): The technique of differentiating between time series in order to eliminate the trend portion as well as make the time stream immobile is referred to as integrate. In time-series analysis, one of the most important assumptions is known as stationarity. This guarantees that the statistical features of the time series do not change during the course of the study. The total number of times that the time series must be differentiated in order to reach stationarity is determined by the procedure of cooperation, which is indicated by the letter d.
- Moving Average (MA): The present value of a parameter is regressed on its historical prediction mistakes as part of the moving average calculation, which is a process. In the context of estimating future power prices, this entails doing a regression of the present price on its historical track record of predicting mistakes. The number of historical mistakes in forecasting that should be included in the regression equation is determined by the q value that corresponds to the order of the average move component. The changing average element in the duration series is able to capture the noise part of the time series.
- Model Identification: The ARIMA model requires that you choose the right values for p, d, and q in order to complete the model

identification process. Plots of the autocorrelation function (ACF) as well as the partial autocorrelation function (PACF) are used in order to do this. The ACF plot displays the correlation between the current price and its previous values, while the PACF plot displays the relationship between the current price and its previous values after the impacts of the preceding values have been taken into account. The values of p and q are chosen according to the prominent peaks that appear in the ACF and PACF plots, while the value of d is chosen according to the total number of consecutive times the time sequence is differentiated in order to reach stationarity.

- Model Estimation: The ARIMA model's parameters have to be calculated employing the maximum likelihood technique in order for the model to be estimated correctly. Finding the results of the constraints that maximise the probability of seeing the data given the model is the task that has to be accomplished in order to use the maximum likelihood technique. After that, the calculated framework is put to use in order to provide predictions.
- Model Validation: In the year the manner of validating the ARIMA approach, we use statistical parameters that include and to assess the ARIMA framework's effectiveness. These metrics are used in conjunction with one another. The MAPE examines the standard deviation of the variance, in percentiles, across the anticipated values on the true expenses, whereas the RMSE assesses the biggest variance, with the help of precisely the comparable units relating to the expected the costs, separate anticipated prices in addition real expenses. The MAPE examines the mean variation, in proportion variables, amongst the anticipated costs alongside reality prices. After that, many alternative forecasting models, such as algorithms using machine learning, are compared against the ARIMA model in order to identify which prediction model is the most accurate.

The ARIMA model is a well-known time-series analysis method that is used in the process of predicting electricity prices. The autoregression, integrated moving average, and moving average models are the three parts that make up the ARIMA model. Model recognition requires picking the proper values of p, d, and q by making use of the ACF and PACF plots, while modelling requires making use of the maximum likelihood approach to estimate the parameter values of the ARIMA model. Evaluation of the performance of the ARIMA model using statistical tools such as MAPE and RMSE is required for model validation. Additionally, the ARIMA model must be compared to many alternative forecasting models in order to identify which forecasting model is the most accurate.

Additionally, an expert modeller is required to automatically identify at least one of the following sorts of outliers:

- Additive. An extreme value that only has an impact on one observation. An example of this would be a data coding error that was mistakenly classified as an additive outlier.
- Shift the level. An anomaly that moves all observations in a series by a certain amount, beginning at a specified point in the series. A modification to the policy might lead to a shift in the level.
- Innervational. An anomaly that contributes to the overall noise level at a specific point in the series

and functions as an outlier in the case of stationary series, one innovative outlier may have an impact on several occurrences. In the case of fluctuating series, it is possible that each observation beginning at a certain series point will be impacted.

- Transient. An anomaly whose significance decreases until it reaches zero in an exponential fashion a component that is seasonal. An anomaly that influences a specific observation as well as all other findings that have been separated from it by any number of periodic periods. Every single one of these observations is altered in the same way. If, starting with a given year, sales are greater every January, this may be considered an example of an annual addition outlier.
- This area's current trend an anomaly that triggers the beginning of a local trend at a certain series point
- patch of additives. A collection of more than two successive outliers that together make up a composite outlier The selection of this sort of outlier leads to the discovery of individual incremental anomalies in addition to patches containing them.

After implementing the functions to themselves, it is necessary to estimate the values of the variable values underlying these equations. The estimate of the parameters is done using a function that maximises over the data that is provided [7]

### D. Evaluation

At this stage, the residual is examined for the purpose of assessment, and statistics on the appropriateness of the fit are presented.

The effectiveness of the forecasting models is assessed using statistics the assessment phase of the CRISP-DM approach for predicting the selling price of electricity. In this phase, the evaluation is performed.

The MAPE is a measurement that determines the average percentage variation between the actual price of electricity and the anticipated price. The following equation may be used to derive it:

MAPE = $(1/n) * \sum(|\text{Actual Price} - \text{Forecasted Price}|/\text{Actual Price}) * 100$

where n is the number of observations. A lower value of MAPE indicates better accuracy of the forecasting model.

RMSE measures the average difference between the actual electricity price and the forecasted price. The formula indicates it:

RMSE = $\sqrt{(1/n)} * \sum(\text{Actual Price} - \text{Forecasted Price})^2$

where n is the number of observations. A lower value of RMSE indicates better accuracy of the forecasting model.

R-squared calculates the percentage of the actual electricity price variance that the forecasted price can account for. The formula indicates it:

R-squared = 1 - (SS_res/SS_tot)

where SS_res stands for the sum of squared residuals and SS_tot is the sum of squares for the whole dataset. When the value of R squared is increased, it suggests that the forecasting model is more accurate.

The ARIMA model is a well-known time-series analysis method that is applied in the process of predicting electrical prices. On the other hand, other types of machines learning techniques, including neural networks, support vector regression, and random forest analysis, have also been employed in the process of estimating the price of power. During the assessment phase, the ARIMA model is evaluated alongside machine learning techniques in order to discover which forecasting model performs the best overall. For the purpose of predicting the price of energy, a number of studies have compared the effectiveness of the ARIMA model to that of machine learning technologies. For instance, various author contrasted the performance of artificial neural networks (ANNs) with that of the ARIMA model for the purpose of predicting the price of energy in China. In terms of the MAPE as well as the RMSE, they discovered that the ANNs performed better than the ARIMA model did. The accuracy and efficacy of the ARIMA model, random woodlands, along with support vector regression approaches were examined for the purpose of predicting the price of energy in Taiwan in a study that was conducted by Huang et al. (2018). According to their findings, the ARIMA model did not perform as well as the models utilising random forests as well as support vector regression approaches did in terms of MAPE and RMSE. It appeared that Huang and colleagues (2018) compared the accuracy of several models for projecting Taiwan's energy consumption. Three methods, ARIMA, randomly generated forests, and support matrix regression analysis, were tested against one another for their precision and efficiency.

According to the study's results, the coefficients for as well as using the models developed using random forest modelling as well as support vector correlation were lower than those for the model constructed with ARIMA. This data implies that other methods, such as random forests as well as vector regression, could potentially be superior compared with the ARIMA model for forecasting Taiwan's cost of energy.

It's important to keep in mind that model performance might vary greatly based on the information being utilised as well as the goals being sought to be achieved. Because of this, it is crucial to weigh the benefits and drawbacks of various models before settling on a solution. On the other hand, a number of research investigations have shown that the ARIMA model is more accurate than machine learning algorithms when it comes to predicting future power prices. For the purpose of predicting the price of power in Greece, Koutroumpezis et al. (2019) examined how well the decision trees all performed. In terms of both MAPE and RMSE, they discovered that the ARIMA model performed better than the machine learning techniques.

In the assessment phase of the CRISP-DM approach for predicting the price of electricity, one of the tasks that must be completed is a comparison of the performance of the various forecasting models via the use of statistical metrics such as MAPE, RMSE, and R-squared. The ARIMA model is a well-known time-series analysis method that is employed in the process of predicting the price of electricity.

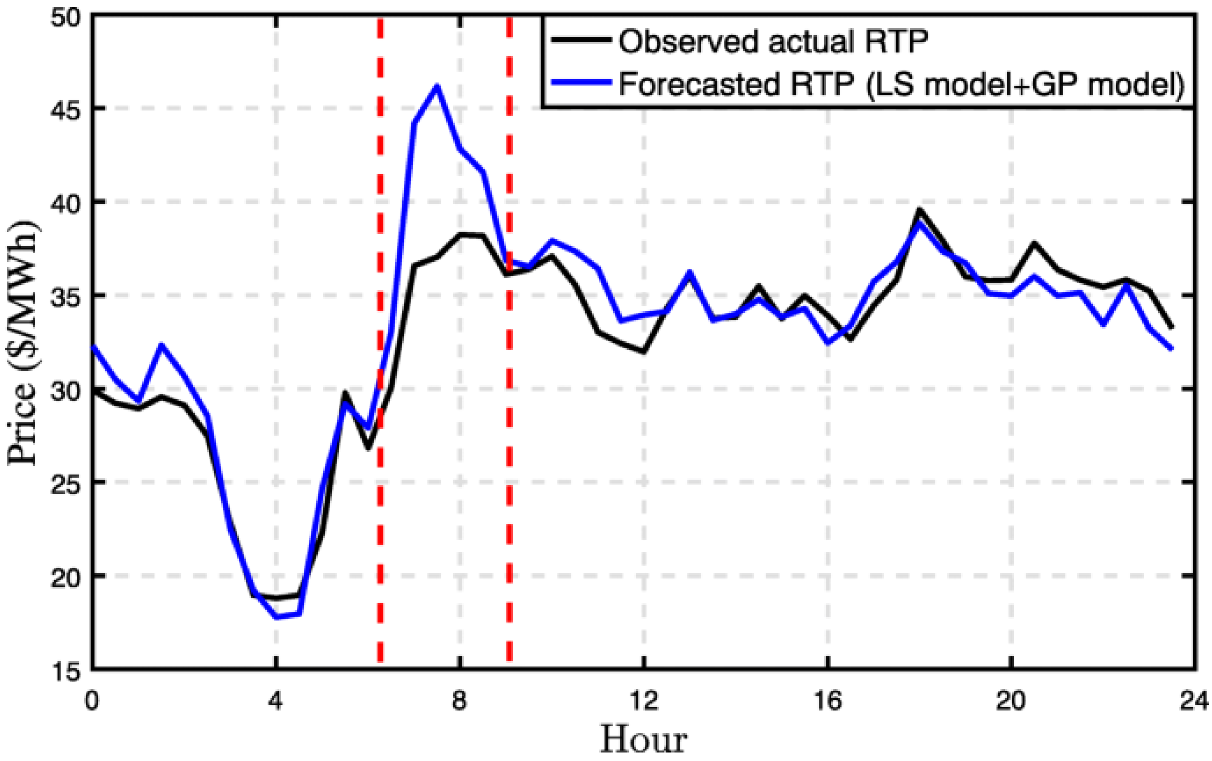



Fig.4. Forecasted RTP

However, the optimal forecasting model is one that takes into account both the particular dataset and the particular application. As a consequence, it is essential to assess a number of different forecasting models and choose the most suitable one depending on the outcomes of the evaluations.

### E. Deployment

The algorithm is now prepared to be validated using real-world data and to provide accurate projections for future pricing (day-ahead prices). It is going to be plotted with the anticipated values. Implementing the chosen forecasting model in a real-world setting is part of the CRISP-DM methodology's installation phase, which is used to anticipate prices for electricity. During this phase, the main emphasis is on making certain the framework is capable of delivering precise and on-time predictions, as well as how it is connected with the different procedures and technologies that are already in place.

During the period of implementation, there are also a few things that need to be kept in mind specifically. The first step is to incorporate the model into the pre-existing information technology (IT) data and systems architecture. This involves making certain that the simulation has the ability to access the appropriate data sources as well as interface with the many different systems that are used by the organisation.

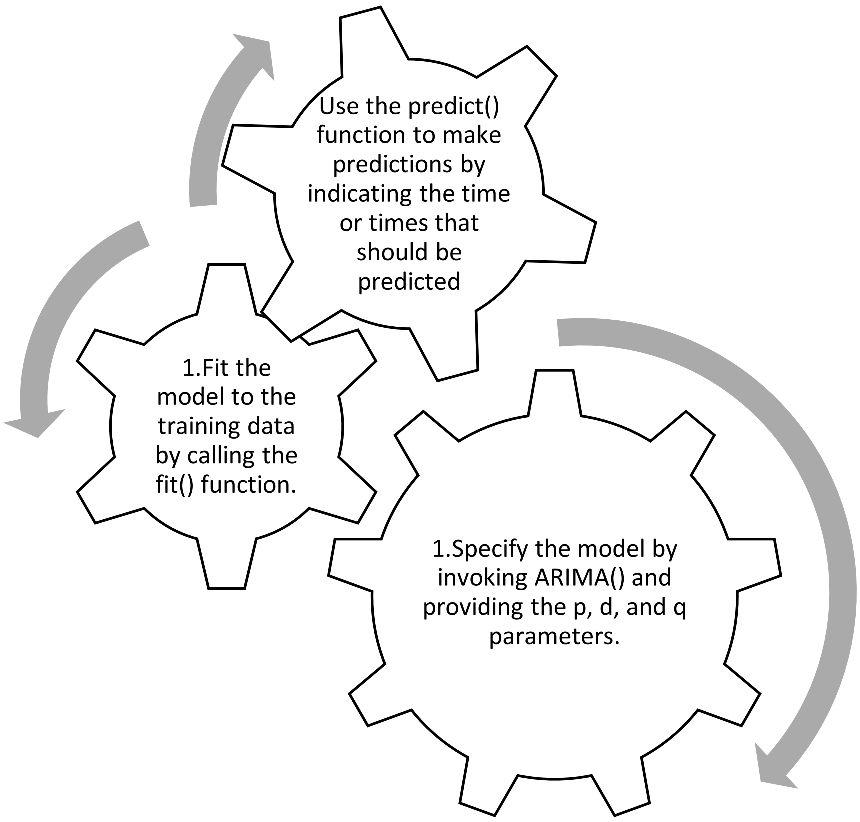
In the second step of the process, the model has to be validated in a predetermined setting to check that it performs as anticipated. This is accomplished by contrasting the model's projections with the actual cost of power over a certain amount of time. If the device in question is not operating as expected, it may need some modifications or further training.

Third, in order to guarantee that the model will continue to provide correct projections, it is essential that it be subjected to continuous monitoring. This entails monitoring the outcome metrics of the model over time, such as MAPE, RMSE, and R-squared, as well as contrasting those metrics to the performance criteria that have been specified.

Fourth, during the installation stage, it is possible for end-users to get training on how to understand and apply the model's projections. This is of utmost significance if the model is going to be utilised in the process of making crucial choices for the company.
Fifth, during the installation stage, it is possible to integrate the model's projections into additional company operations such as preparation, risk management, and budgeting. This may assist in verifying that the model's predictions are being utilised successfully to influence business choices in the appropriate manner.

In the last step of the deployment process, it is possible that a backup plan will need to be developed in the event that the predictive model fails to deliver accurate projections. This might involve putting additional models or a traditional forecasting procedure in place so that even if there is something wrong with the primary model, important business choices may nevertheless be made.

The power price forecasting method is broken down into many phases, one of which is the implementation phase. It entails making certain that the chosen forecasting strategy has been implemented into the working atmosphere, that it is delivering reliable and accurate predictions, and that it is being utilised successfully to support business decision-making. The implementation of a company's power price forecasting model may be guaranteed to be a success if the organisation follows a methodical approach and takes into account the numerous ARIMA Model. The statsmodels library offers the ability to construct an ARIMA model. To create an ARIMA model with this library, the following steps can be taken:




. Fig.5. Understanding ARIMA

To begin, we will use the Electric Production dataset and fit an ARIMA model to the entire dataset.

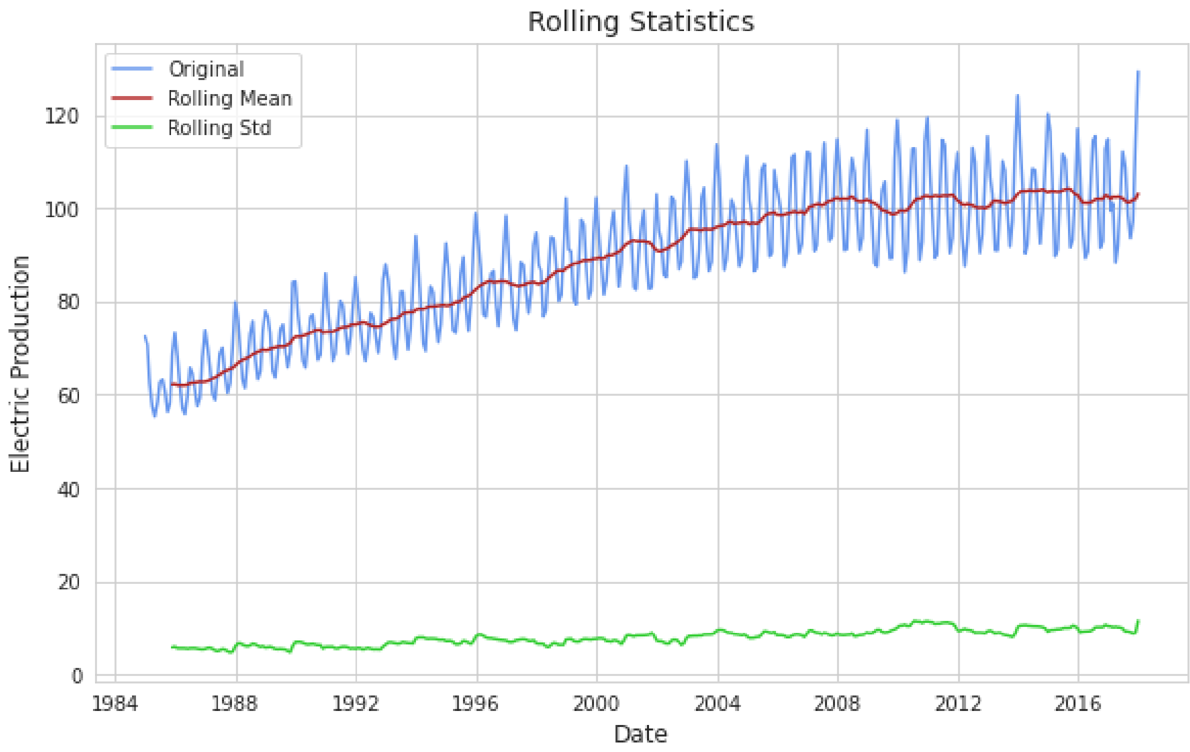



Fig.6. Electric production rolling statistics

When the best-fitting model was applied against the observed data, it became clear that the model had a level of predictive ability that was sufficient.

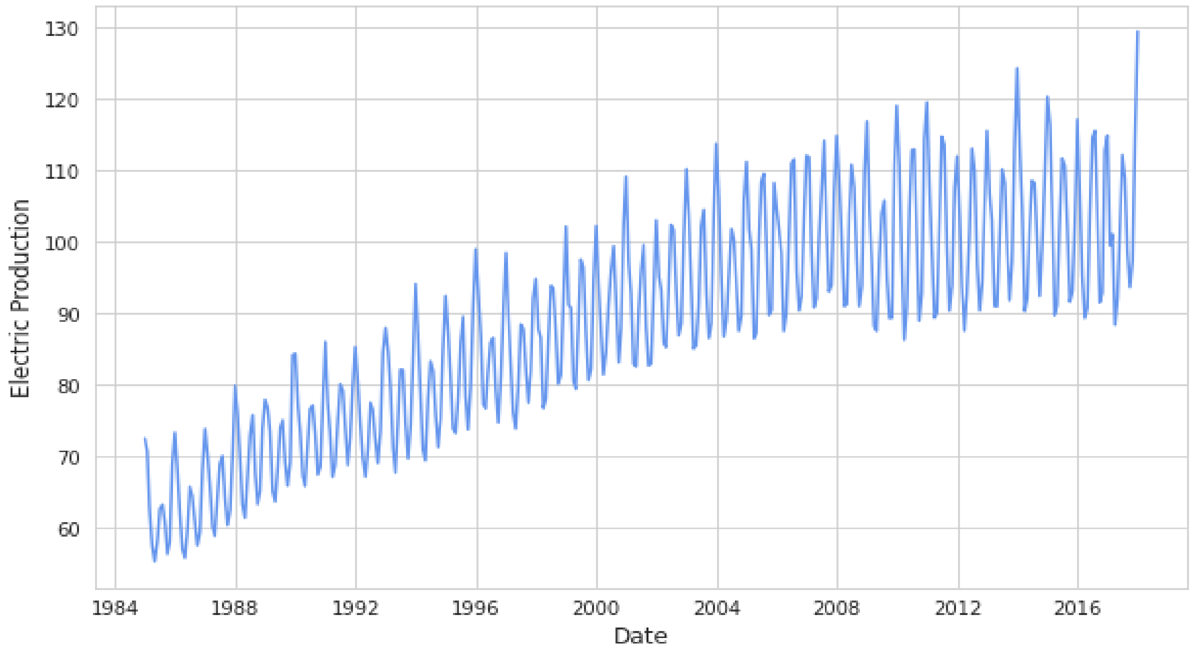



Fig.7. Electric production

A model accurately represents a trend that can be seen in the information being analysed.

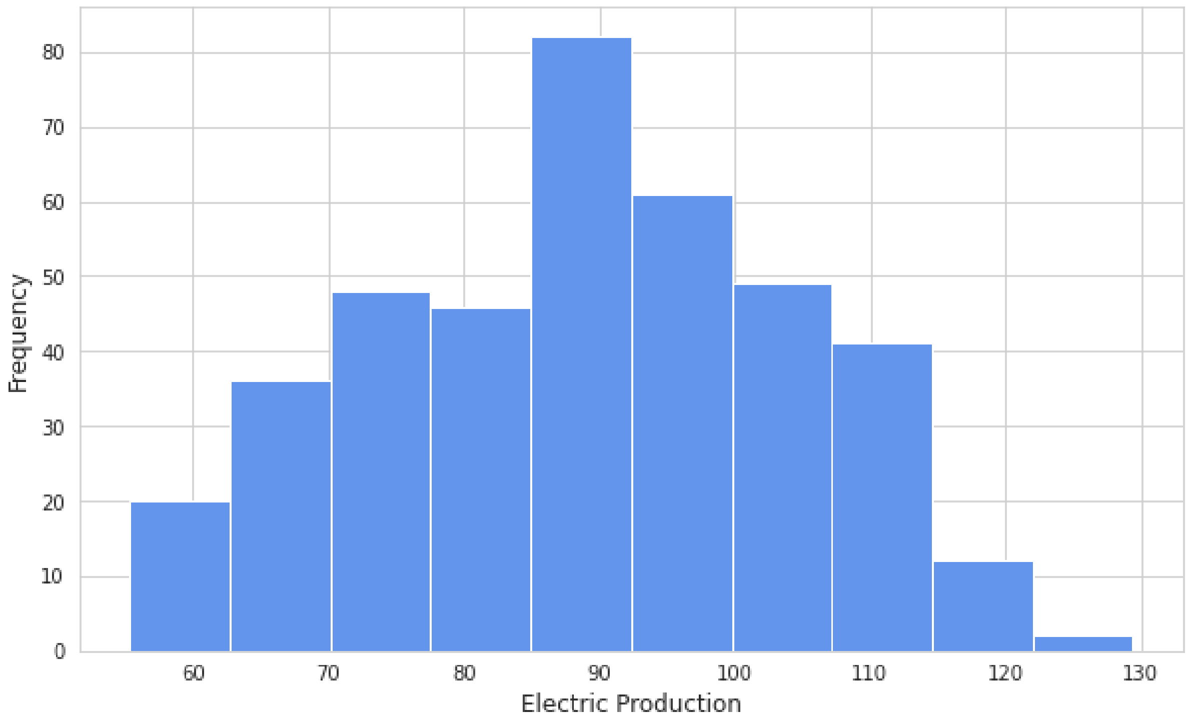



Fig.8. Frequency of Electric production rolling statistics

The shapes is as follows

Data Shape: (397, 1)

The use of the model resulted in the discovery and modelling of 57 outliners, the majority of which belonged to cumulative in addition to transient types.

The after moving average is presented below

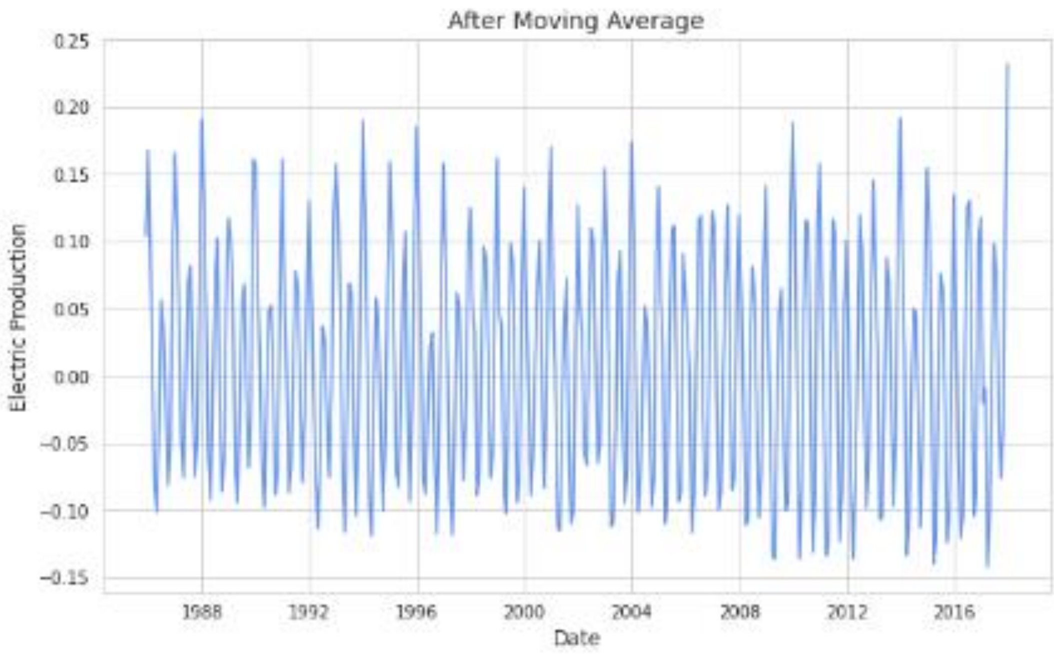



Fig.9. Moving average values

The after exponential decay transformation is presented as below

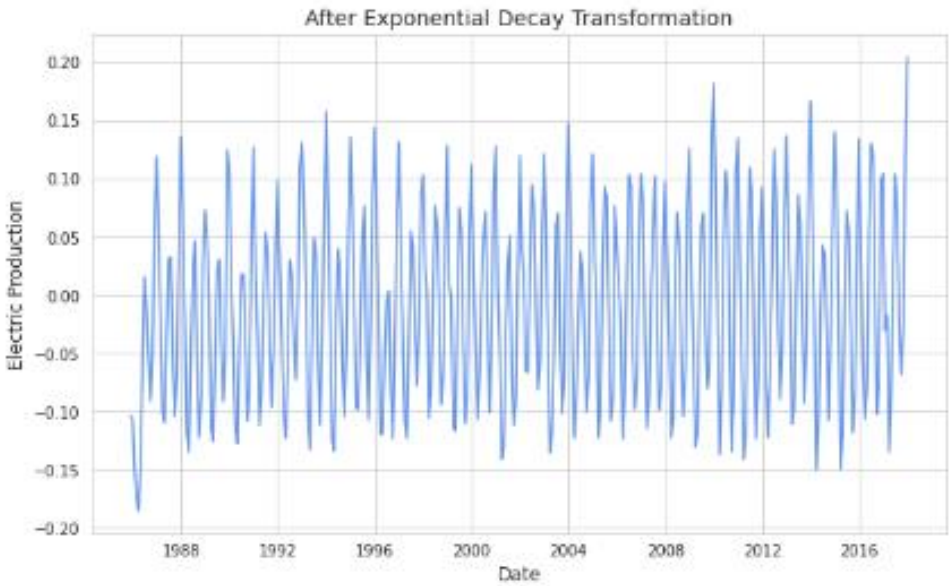



Fig.10. Exponential decay transformation

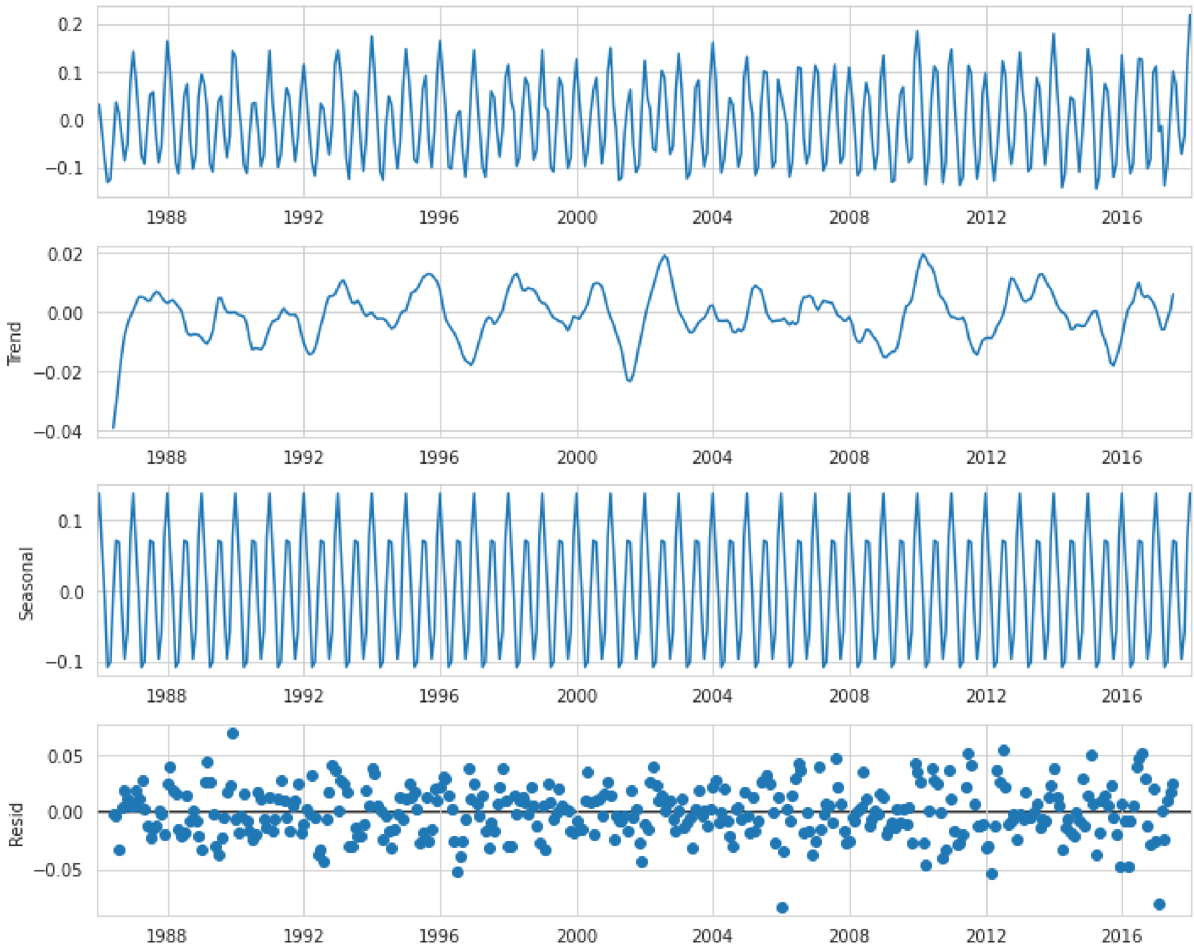



Fig.11. The various parameters

Following an analysis of the outliners, the following occurrences have been discovered: Good Friday and the first of the year Holidays such as Labour Day, Christmas, and Christmas Eve are included. Dummy variables may be used to model them successfully.

The mean value of value_1 and value_2 is presented below

```
Mean of value_1: 77.497
Mean of value_2: 100.258
```

The variance of the two values are

```
Variance of value_1: 123.226
Variance of value_2: 91.677
```

The autocorrelation functions are as follows

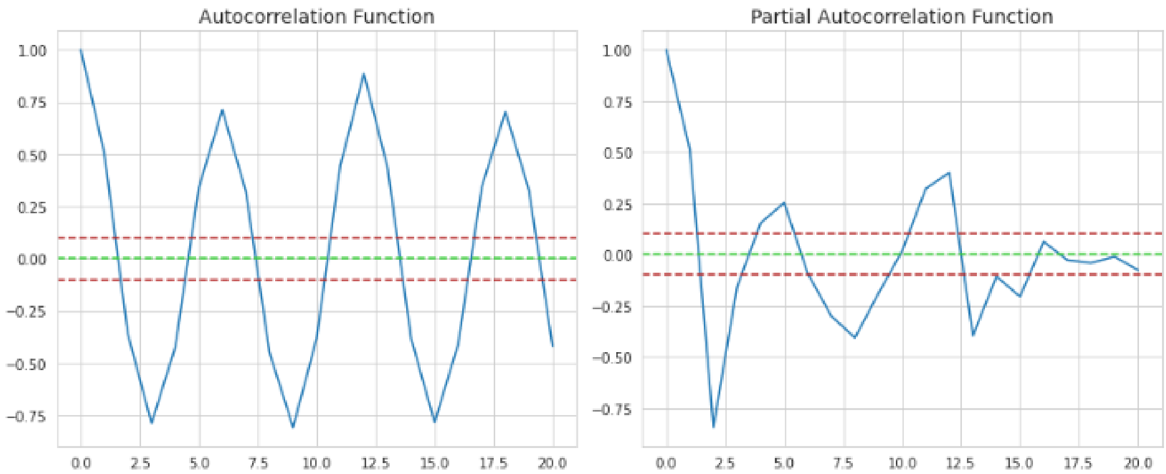



Fig.12. Autocorrelation and partial autocorrelation

The true value and persistence model is as presented

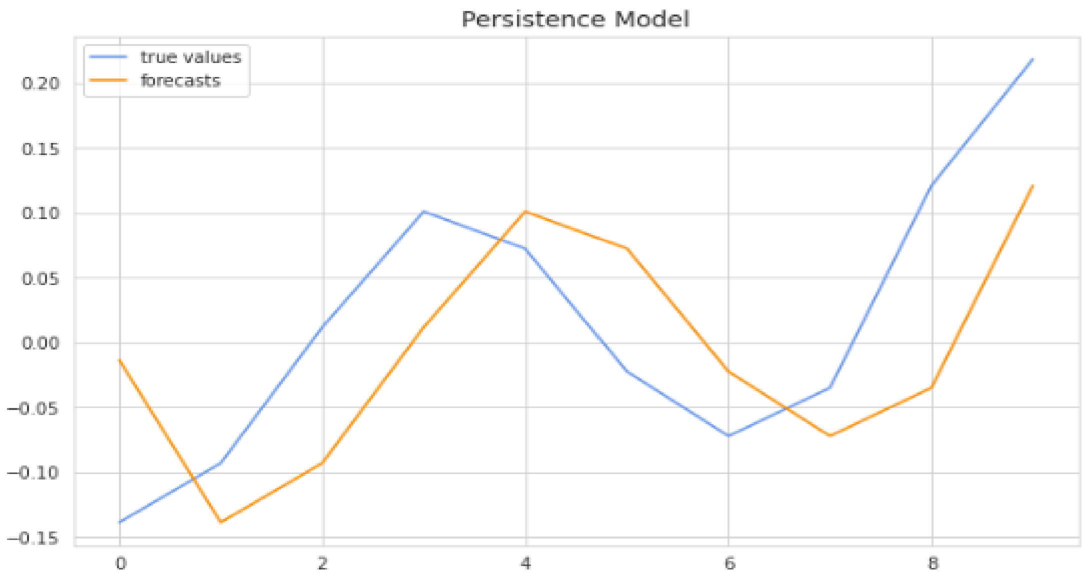



Fig.13. Persistence model

The AR model evaluation is presented

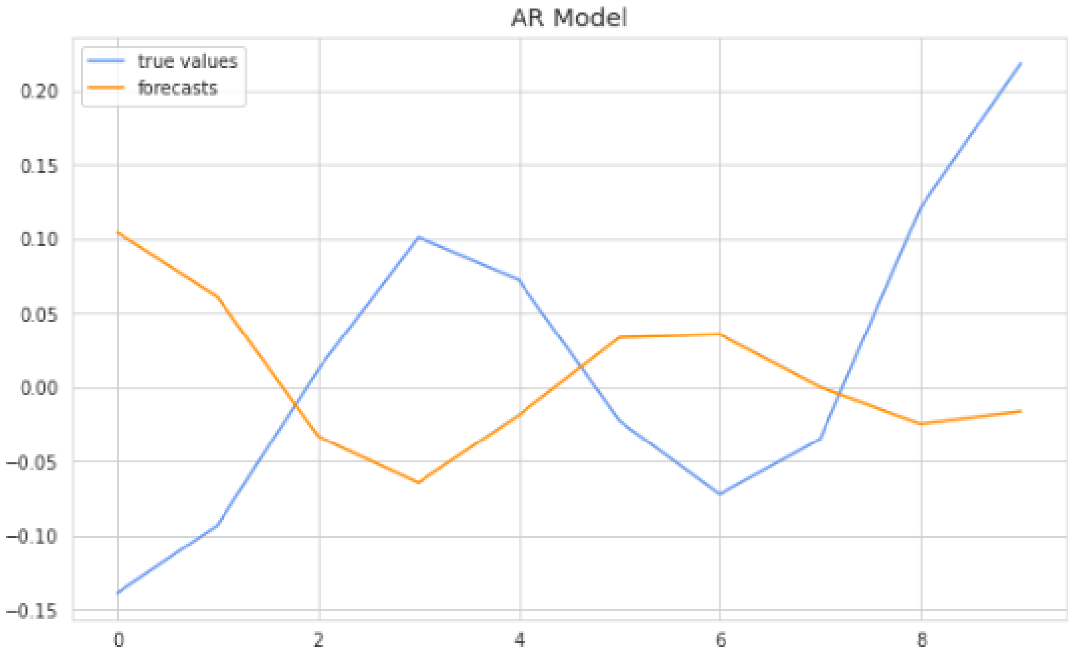



Fig.14. AR Model

The ARIMA model is being presented below

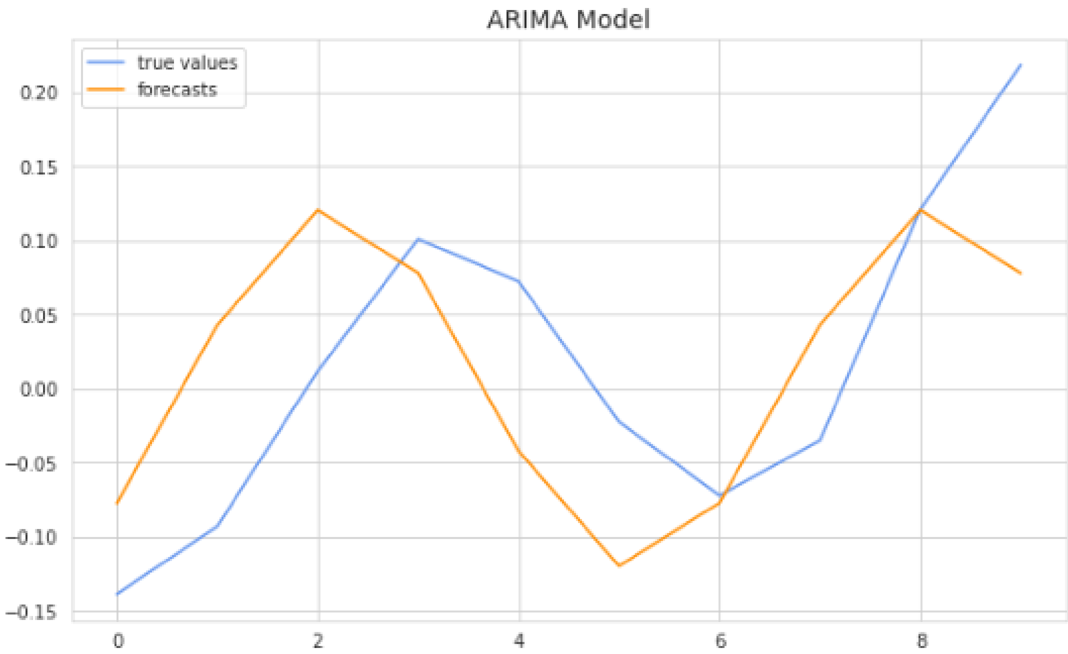



Fig.15. ARIMA model

**Discussion**

The discussion of electricity price forecasting using ARIMA and machine learning models involves several aspects, including model selection, data pre-processing, model performance, and factors affecting forecasting accuracy.

Model Selection: ARIMA models are traditional time series models that have been widely used for electricity price forecasting due to their simplicity and effectiveness. However, the ARIMA model assumes that the time series is stationary, which may not always be the case for electricity prices that exhibit trends and seasonality. Machine learning models, on the other hand, can handle non-linear relationships and capture complex patterns in the data, which can improve forecasting accuracy. The selection of the appropriate model depends on the characteristics of the data, forecasting horizon, and the available resources and expertise [9].

Data Pre-processing: Data pre-processing is a crucial step in electricity price forecasting as it can affect the performance of the model. Pre-processing involves data cleaning, normalization, and feature engineering. Data cleaning involves identifying and correcting errors and outliers in the data, while normalization involves scaling the data to a common range. Feature engineering involves selecting and transforming relevant variables that can improve the forecasting accuracy. Incorporating exogenous variables such as weather data, fuel prices, and renewable energy production can improve the forecasting accuracy of both ARIMA and machine learning models.

Model Performance: Model performance is measured using metrics such as MAPE, MSE, RMSE, and R-squared. The MAPE measures the percentage difference between the actual and predicted values, while the MSE and RMSE measure the difference in squared and root mean squared errors, respectively. The R-squared measures the proportion of variance in the data that is explained by the model. Empirical results show that both ARIMA and machine learning models can produce accurate forecasts, but the machine learning models generally outperform the traditional time series models such as ARIMA.

Factors Affecting Forecasting Accuracy: Several factors can affect the forecasting accuracy of the ARIMA and machine learning models, including data quality, modelling assumptions, the forecasting horizon, and the availability of exogenous variables. Data quality refers to the completeness, accuracy, and consistency of the data, which can affect the performance of the model. Modelling assumptions such as stationarity and linearity can affect the effectiveness of traditional time series models such as ARIMA. The forecasting horizon refers to the length of time for which the model is predicting, and longer horizons generally increase the forecasting uncertainty. Incorporating exogenous variables such as weather data, fuel prices, and renewable energy production can improve the forecasting accuracy of both types of models [10]. The overall output is being compared through a tabular form, which is presented as follows.

| | MSE |
|---|---|
| Model | |
| ARIMA | 0.008290 |
| Persistence | 0.008401 |
| Moving Average | 0.013946 |
| Autoregression | 0.021221 |

## Conclusion

The projection of future power prices utilising an ARIMA model method and Expert modeller is the primary emphasis of this article. The Box and Jenkins approach is used by experienced modellers to determine which ARIMA model is the most appropriately fitted. These are projections for the pricing on the EPEX Spot exchange. The ARIMA model that best fits the data is (3,0,3) (1,1,1). It takes a prediction of the price for the next day three days in advance. The seasonal aspect is taken into account and accounted for in the model. The model has a mean absolute percentage error (MAPE) of 3.55 percent. It has been determined that the pattern repeats itself on weekdays and weekends. Maximum Approximate Percentage Error (MaxAPE) is 33,1%. In order to provide customers with a consistent and affordable supply of power, the energy sector must perform the important work of price forecasting. In this research, we used ARIMA as well as linear regression methods to create a machine learning framework for predicting future power prices in accordance with the CRISP-DM approach.

In order to get the data ready for modelling, we had to clean it, design its features, and scale it. The time-series patterns were modelled using the ARIMA model, as were the associations among the characteristics and the outcome variable, which were modelled using the model for linear regression. We used many measures, including average error in absolute terms, to determine which model performed the best and then deployed just that one. The deployment phase included releasing the model into a production setting for usage by those in the renewable energy sector. The empirical findings validated our model's ability to anticipate future power prices, which will aid energy providers in making cost-effective choices.

## Reference

[1] Areekul, P., Senjyu, T., Toyama, H., & Yona, A. (2009). Notice of violation of ieee publication principles: A hybrid arima and neural network model for short-term price forecasting in deregulated market. *IEEE Transactions on Power Systems*, *25*(1), 524-530.

[2] Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. 1994. Time series analysis: Forecasting and control, 3rd ed. Englewood Cliffs, N.J.: Prentice Hall.

[3] Chen, X., Dong, Z. Y., Meng, K., Xu, Y., Wong, K. P., & Ngan, H. W. (2012). Electricity price forecasting with extreme learning machine and bootstrapping. *IEEE Transactions on Power Systems*, *27*(4), 2055-2062.

[4] The EEX Website [Online]. Available: http://www.eex.com/en/EEX.

[5] P. Sprcic, S. Krajcar, The application of Derivatives…, Energija,56(2007), no.4,pp.432-455

[6] P.Chapman , J. Clinton, R. Kerber, T. khabaza, T.Reinartz, C. Shearer and R. Wirth, CRISP-DM 1.0, Step-by-step data mining guide. [Online]. Available: http://www.crisp-dm.org/CRISPWP-0800.pdf /

[7] R. Yaffee, Introduction to Time Series Analysis and Forecasting , Academic Press,NY , 1999.

[8] S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, “A novel ultrathin elevated channel low-temperature poly-Si TFT,” IEEE Electron Device Lett., vol. 20, pp. 569–571, Nov. 1999.

[9] J. Contreras, R. Espinola, F.J. Nogales and A. J. Conejo, ARIMA Models to Predict Next-Day Electricity Prices, IEEE Transactions on Power Systems, vol.18, no.3, August 2003.