COMPARATIVE ANALYSES OF FORECASTING TECHNIQUES FOR ELECTRICITY WHOLESALE PRICE UNDER HIGH PENETRATION OF RENEWABLE ENERGY SYSTEMS

*Joseph Daly[1],Lin Zheng[1], Min Xuan[1], Yunxin Yang[1], Mattia De Rosa [2],Fabiano Pallonetto[1]**

[1]*School of Business at Maynooth University,Maynooth, Ireland*
[2]*University of Sassari, Sassari, Italy*
**E-mail: fabiano.pallonetto@mu.ie*

Keywords: ENERGY, FORECASTING, ELECTRICITY PRICE, RENEWABLE, MACHINE LEARNING

Abstract

Optimising electricity production and consumption requires fuel suppliers, electrical grid operators, trading operators and demand response aggregators to monitor and predict the wholesale electricity price. It is also essential for integrating renewable energies and dispatching electricity generators. Using predictive models to forecast wholesale electricity prices is highly beneficial to these businesses to prepare for cases where the price deviates from normal patterns. This report focuses on predicting the wholesale electricity price in Ireland using cleaned data from 2015 to 2021, containing specific features. With a literature review of the related work in this field, two models were selected to analyse the dataset an ARIMA and ETS. These concepts are explained, and the results of the models are discussed for forecasting 14 days and 30 days. The predictive accuracy of these models was compared by using the evaluation metrics MAPE, MAE and RMSE. The results show how the ETS model best predicted wholesale electricity prices in both time horizons. Although there are limitations associated with the data and predictive models, this paper highlights the potentiality of these predictive models in wholesale electricity price forecasting and envisions future steps to improve their performance.

1 Introduction

Environmental impact, political instability, increased penetration of renewable energies, the continued growth in energy demand, and the uncertainty of fuel reserves are new challenges that the power systems research community is addressing. Grid reinforcement and more generation can be part of the solution to these challenges; however, it is costly and does not always improve the system's robustness. Recent blackouts in Germany, Texas and Italy caused by a domino effect of small evaluation mistakes are the empirical evidence of a more significant research problem [9, 15]. Addressing these issues requires complex modelling and extensive computational capabilities and advanced computational techniques. To correctly dispatch generators and embed more renewable in the system, the forecasting of wholesale electricity prices is one of the most important areas that many organizations and researchers focused on. Accurate price forecasting is critical for stakeholders of the market to determine their ideal bidding tactics, manage associated risks, and improve asset management.

To gain a deeper understanding of this sector, the current paper focuses on predicting the wholesale electricity price in Ireland. Firstly, a literature review will be adopted to give a general idea of predicting electricity price, as well as offers academic support for the methods we would use later. Secondly, we chose to use and compare two prediction models: ARIMA and ETS. The logic and methodology behind each step of these models will be introduced. The previous consultancy project provided a great fundamental to this study, we will continuously dive into the energy market in Ireland to compare the accuracy of different electricity price forecasting methods.

2 Related Work

In this section we are going to analyse the different techniques utilised for the forecasting of the electricity price. ARIMA model is one type of Time Series analysis. ARIMA represents Auto Regressive Integrated Moving Average, which has been used widely to forecast commodity prices, such as stocks, oil, and natural gas [10]. Numerous case studies demonstrated the efficacy of utilizing ARIMA to forecast the price of electricity. For instance, the ARIMA model was successfully applied to forecast power prices in the Spanish mainland and the Californian markets, respectively, and have shown promising results [3]. Moreover, Jakaa et al. [12] concluded that the model has a reasonable predictive capacity, as it catches the trend of the data well and predicts seasonal peaks. [14] utilized ARIMA to estimate daily electricity day-ahead prices for the German energy market and concluded that the model has a reasonable predictive capacity, as it catches the trend of the data well and predicts seasonal peaks.

The wholesale electricity price dataset from 2015 to 2021 in Ireland is daily series and shows a typical seasonal feature. In addition, the iterative process of ARIMA as well as it allows us to add more covariates to train the model, which would offer us more options when training models. Therefore, we argue that ARIMA may be a suitable prediction model to support TSOs in modelling price prediction trends and forecast the next few day's electricity prices.

ETS is a commonly used one-step supervised forecasting model on univariate datasets based on an extended exponential smoothing method, consisting of a level component, a trend component (T), a seasonal component (S), and an error term (E). ETS model as one of the widely used time series models is successfully applied for electricity price prediction and showed promising performance in terms of accuracy for shore-term(daily) forecasting. In the research on Spanish electricity market price prediction[6], the exponential smoothing model for double seasonality presented the best general result in the group of univariates models, specifically suitable for the longer forecasting horizons than ARIMA. ETS models also provided smaller errors and thus better accuracy in one India electricity price prediction compare with different univariate forecasting models[14]. For Electricity Price Forecasting for Nord Pool Data, Exponential smoothing made the most precise prognosis, with MAPE (Mean Absolute Percentage Error) being only 1.25% [13].

2.1 Critical Evaluation

During the literature review, we discovered that the two most common approaches for forecasting electricity prices are machine learning and statistical models. A general limitation of these models is that the performance of the model decreases significantly in predicting extreme values. For example, an ARIMA model was used to predict the highest air temperatures in Dublin. The results show that more than 70% of the predicted values deviate from the true value within 10%. However, the range of predicted values is from 4.8°C to 28.7°C, which cannot be viewed as extremely temperature in Dublin [1]. It is because ARIMA model has limitations when it comes to forecasting extreme values. It has better performance on modelling seasonality and trends, but it is difficult to forecast unusual values which fall outside of the general trend represented by the model [2]. To solve this problem, some researchers discard the extreme values as outliers, and some researchers classified the target variable into normal and unusual groups to predict. Moreover, it is usual to combine different models together in the real world. For example, we can first use a decision tree to classify the target variable, then use Time Series to predict values in different classification to reach a better result. Or Time Series model can be used to predict continuous numeric values which has been used as a predictor in Regression Tree. How to use these models is dependent on the specific context of each study. For our study, the main objective is prediction of wholesale electricity price by using different feasible models, and to figure out which model will provide a better result with a structured comparison. Thus, we will remove the extreme values as outliers, and focus on the predictor's identification and model comparison.

3 Methodology

This study followed a clear workflow as shown in the below chart. Firstly, the objective has been defined with proper motivation and research question as shown in the introduction section. Secondly, the data collection contains two parts: (1) a dataset with Penetration of Renewables, Wholesale Electricity Price, Total Demand, Wind Available, Wind Generation, and Total Renewables Generation; (2) more records of potential independent variables have been collected from Energy Chart [3], data includes daily generation from Hydro Run-of-River, Fossil Hard Coal, Fossil Oil, Fossil Peat, Fossil Gas, Wind Onshore. Energy Chart is an integrated resource website operated by Fraunhofer Institute for Solar Energy Systems ISE, which can be viewed as a reliable source. The raw data is from 2015-01-01 to 2021-12-31 due to availability.

Model name	Forecasting	Data type	Time range	Days
ETS	14-days	Training data	20.6.2020 - 30.05.2021	275
		Test data	31.5.2021 - 21.6.2021	14
	30-days	Training data	20.6.2020 - 30.04.2021	259
		Test data	21.4.2021 - 21.6.2021	30
ARIMA	14-days	Training data	20.6.2020 - 30.05.2021	275
		Test data	31.5.2021 - 21.6.2021	14
	30-days	Training data	20.6.2020 - 30.04.2021	259
		Test data	21.4.2021 - 21.6.2021	30

Table 1 Parameter setting of ETS and ARIMA models.

Thirdly, in the data cleaning step, the data has been converted to daily at the beginning. A standard format of the cleaned dataset was used to aggregate all data. Then, the Scatter Plot and Interquartile Rule have been applied to find and remove the outliers. The criteria of outliers are 1.5 times IQR (Q3 – Q1), which means the range of non-outlier records is from Q1 – 1.5*(IQR) to Q3 + 1.5*(IQR). Fourthly, we used correlation test and descriptive statistics to find the highly relevant independent variables and useful insights from the data. The cleaned dataset contains 2248 rows and 13 columns, the last column 'Wholesale Price' is the target variable.

From the data quality report, we can see the standard deviation of wind-related values is higher than non-renewables'. This shows a major problem with using renewable energy: the unstable of renewables. Wind energy provides the largest share of electricity generation, however, the unpredictable of wind brings many challenges to generation and price forecasting. From the total demand and total generation records, we can see that generally the generation is higher than the demand. There are 862 records that the demand is higher than supply, which has an average 27% Penetration Rate and 69,757 MWh of Wind Availability. For the days that supply is higher than demand, the average Penetration Rate is 31% and Wind Availability is 112,976 MWh. Thus, wind energy can be viewed as a significant variable of Total Generation.

The correlation test and linear regression model have been used to find the significant variables. There are five variables that we decided to use as the predictors: Hydro Run-of-River, Fossil Oil, Total Demand, Wind Availability, and SNSP. Fifthly,

Model	Horizon	ME	RMSE	MAE	MPE	MAPE	MASE	ACFI
ARIMA	14-days	1.2512	11.2180	8.3453	-2.8395	18.0778	0.6332	-0.0269
	30-days	1.2307	11.0647	8.2562	-2.9353	18.2972	0.6295	-0.0146
ETS	14-days	0.5511	11.5853	8.6679	-4.2185	18.8741	0.6577	0.1130
	30-days	0.5905	11.4436	8.5648	-4.2690	19.0896	0.6530	0.1380

Table 2 Caption

two models have been developed to predict the wholesale electricity price in this step. After reviewing the literature, we decided to use ARIMA and ETS models. We used a software toolbox to fit the models and tried different combinations of variables to reach a better result. The results will be shown in the next section. Sixthly, the model comparison is a crucial step in this project. Besides the model reports functions, we calculate the MAPE, MAE, and RMSE of the predicted values and actual values. The performance of these three models would be compared to find the advantages and disadvantages section.

4 Results

4.1 ARIMA model

When it comes to time series analysis, ARIMA processes are a class of stochastic processes that can be exploited. Box and Jenkins are responsible for the development and use of the ARIMA methodology for the study of time series analysis [8]. The general ARIMA method is formulated as follows, where Pt is the price at time t, and are functions of the backshift operator B: B1 p=pt1 and is the error term.

ARIMA concludes three components, AR, I, and MA, which are characterized by p, d, q. “AR” refers to Auto Regressive, it is a linear regression model that uses its own lags as predictors. “I” means integrated, Because ARIMA models are stationary, if there is any evidence of a trend or seasonality in the data, it must be transformed through a process called differentiating. “MA” refers to moving average, it demonstrates that the variable of interest is linearly dependent on its current and previous values. In other words, in the forecast model, all previous observations are given the same amount of weight.

To fit a better model, we built some ARIMA models by using different sizes and years of training data and test data, such as we used five years of data (2015-2018) to predict the two years (2019 –2021) electricity prices, used one-year data ranging from 2020 to 2021 to predict 14-days and 30-days electricity prices, and we also add some covariates that have a high relationship with electricity prices to compare the performance of models that without adding covariates. these covariates are Hydro Run-of-Rive, Fossil Oil, Total Demand, Wind Availability, SNPN. Finally, we found the models with covariates get better performance. Below will present the results of 14-days and 30-days ARIMA models with covariates. All the outliers have been removed. (The method shown in the methodology).

For the 14-days forecasting, ME, RMSE, MAE, MAPE, MASE, and ACF1 are calculated. The error measures of the model with covariates are smaller than the model without covariates. The forecast plot (figure 12) shows the historical electricity price data in gray and the expected data in blue.

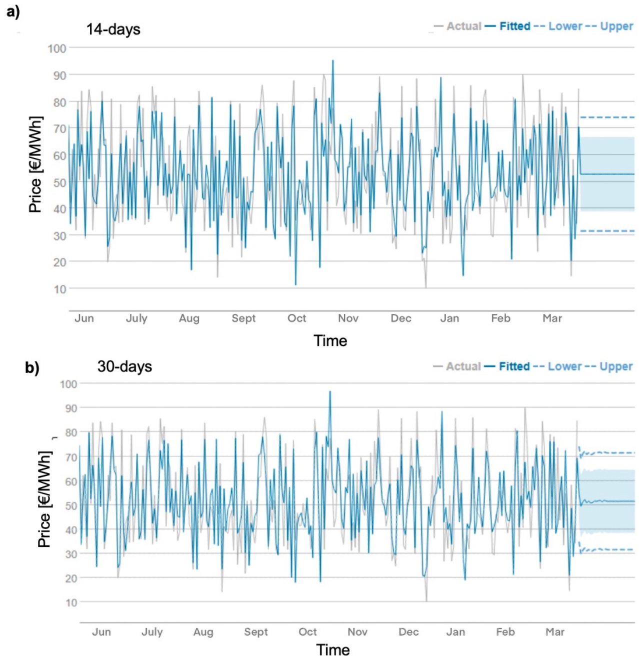

Fig. 1 ARIMA model training for 14-days and 30-days forecasting.

This model predicted the electricity trends correctly. The area between the blue dashed line in the plot shows the 95% confidence interval, and the blue area shows the 80% confidence interval. Similar results of 30-days forecasting to the 14-day forecasting, the performance of the model with covariates is better than the model without covariates. And there is no big difference based on the forecasting results between 14-days and 30-days forecasting models.

4.2 ETS

Exponential smoothing is a historical forecast method proposed in the later 1950s, which had been widely used in industries. Considering the information value of the recent data may be higher than older data, the exponential smoothing assigns a set of exponentially declining weights to past data. For providing better performance which aligns with distinctive features of data, there are many types of the exponential method according to the combination of error, trend, and seasonal components[17].

According to the different components, Figure 17 presents the equations for all the models in the ETS framework[17]. For the hyperparameter setting, because the tool “ETS” developed by Alteryx can use fully automated methods to model the three

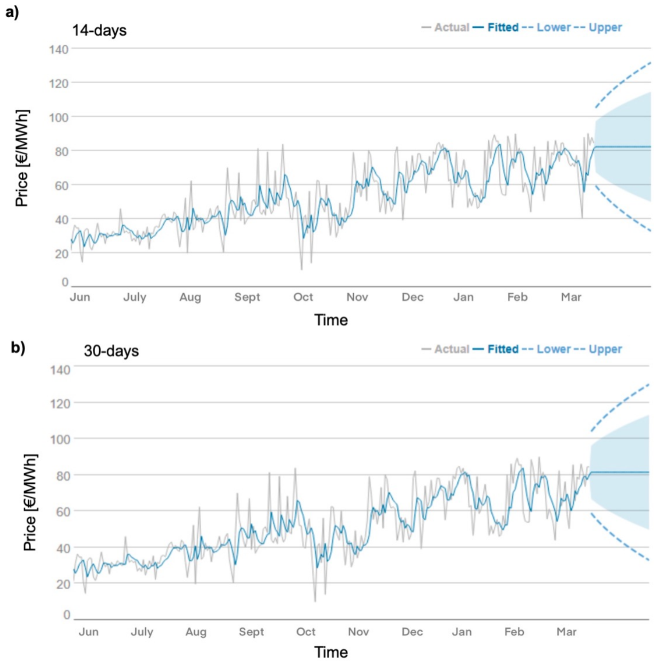

Fig. 2 Results from TES model for 14-days and 30-days forecasting.

components in the "best way" based on statistical criteria, all the hyperparameters were set as “auto”.

For both models, the fitting algorithm is ETS (A, N, N), which means the simple exponential smoothing with additive errors. Both the forecast of 14 days and 30 days plots showed that the fitted value generally catches the trends as the actual value. Besides, the confidence interval of 95% (areas between the dashed blue line) and 90% (areas of blue shaded) become wide with the time close to the end of the forecasting time arrange, which indicated the error of forecasting are larger in the longer-term projections.

The same cleaned dataset as other method which including the data from Jan. 2015 to June 2021 were introduced for modeling. As we learned that ETS model are better performed when use the recent data to predict the shore-term electricity price. The sub-dataset of 1 year from 2020.6.20 to 2021.6.21 were selected for analytics. We set two different forecasting periods, which are the latest 14 days and 30 days of this selected dataset as the test data to compare the performance of varies forecasting horizons.

The errors of these two models are measured in the table below. The MPE (mean percentage error) of both models is negative, which indicated the values of the forecasting are lower than the actual values. Moreover, the performance of the ETS-14 is slightly better than the ETS-30. The errors of these two models are measured in the table below. The MPE (mean percentage error) of both models is negative, which indicated the values of the forecasting are lower than the actual values. Moreover, the performance of the ETS-14 is slightly better than the ETS-30.

Type	Model	MAPE	MAE	RMSE
14-days forecasting				
Training	ARIMA	0.185	7.799	10.729
	ETS	0.189	8.668	11.590
Testing	ARIMA	0.287	24.200	26.010
	ETS	0.052	4.360	4.670
30-days forecasting				
Training	ARIMA	0.171	7.927	9.832
	ETS	0.191	8.565	11.444
Testing	ARIMA	0.324	25.590	27.500
	ETS	0.130	8.267	11.756

Table 3 MAPE, MAE and RMSE values

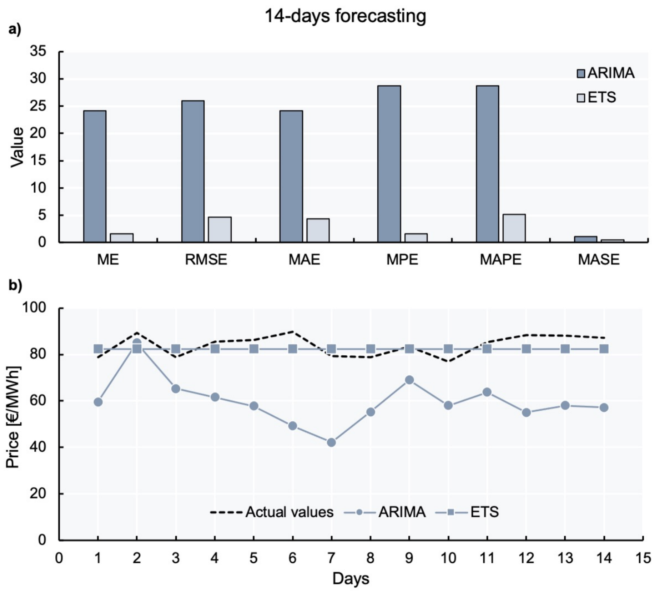

Fig. 3 Comparison between the ARIMA and ETS results for a 14-days forecasting.

5 Comparison of the models

To compare the three models for 14-day and 30-day price prediction, both Jain and Mallick[17] and Pallonetto [5] suggest the use of MAE (Mean Absolute Error), RMSE (Root Mean Squared Error) and MAPE (Mean Absolute Prediction Error) as evaluation metrics. These metrics are used to measure prediction error and can affect the performance of prediction models. For example, MAPE represents the average absolute percent error. Hence, it is important to minimize these values.

From looking at the MAPE, MAE and RMSE values below for the training datasets, they are not significantly different between the 2 models. However, using the testing dataset, the values for the ETS model are far better (smaller) than ARIMA. This is the case for both 14-day and 30-day price forecast. From these results, it is clear that the ETS model is best for predicting the wholesale electricity price in Ireland.

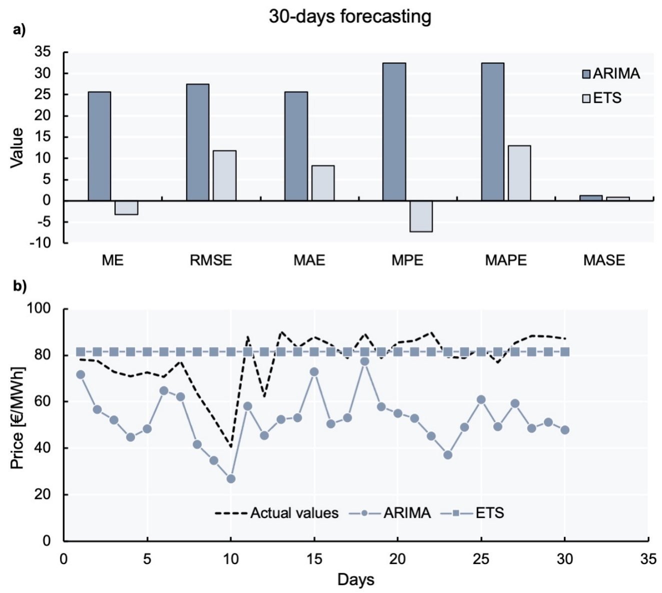

Fig. 4 Comparison between the ARIMA and ETS results for a 30-days forecasting.

6 Conclusion

Through the research of literature reviews, we discovered that both ARIMA and ETS, both time series models, are all suitable in predicting wholesale electricity price. Using clean data from 2015-01-01 to 2021-12-3 consisting of features such as total electricity demand and wind generation in Ireland, we identified ETS as the best model in forecasting wholesale electricity price for 14 and 30 days in the country due to the model's low MAPE, MAE and RMSE values.

Renewable penetration rate has a small negative impact on electricity wholesale price however the relationship can be highlighted by a basic multiple linear regression model and impact on the seasonality of the price. Because of the political instability and the fuel price dependency, the relative error in short term versus long term electricity wholesale prices forecasting is different order of magnitude. Also weather affect significantly the prediction For example, if the predicted wholesale electricity price is high for a certain day, this could be caused by weather conditions that impacted the electricity demand. Such models could be also incredibly useful for suppliers in buying when prices are low and avoiding when they are high.

There are several limitations associated with the research and analysis carried out, the first involving the dataset itself. As the dataset contains daily values from 2015 to 2021, there are only 2,247 rows in the dataset. This brings issues such as the effect outliers have on the performance of the model and overfitting. In future work, we would like to collect data from more previous years, use hourly values instead of daily and include extra variables such as temperature. This could be achieved by using a different source of data or combining different sources. The inclusion of these would greatly increase the size of the dataset and allow us to produce more accurate models. We are also interested in including data from 2022 to forecast wholesale electricity prices for May 2022 to appreciate the true application of these models. The other main limitation is associated with the models themselves. For example, machine learning models are sensitive to the dataset outliers and are likely to overfit the data if there are a lot of features. For future works, Pallonetto[5] discusses the use of LSTM (Long Short-term Memory Network) in forecasting electricity demand. In certain cases, the MAPE value was reaching as low as 2.57% so although we are predicting price, it would be object of future work to see what kind of results would be achieved with models such as LSTM.

7 Acknowledgements

This work has been funded by Science Foundation Ireland, project NexSys SFI Grant No.: 21/SPP/3756

8 References

Websites

[1] 'SARIMA: Forecasting seasonal data with Python and R,', https://medium.com/analytics-vidhya/sarima-forecasting-seasonal-data-with-python-and-r-2e7472dfad83 , accessed 27 May 2022

[2] "Limitations of ARIMA: Dealing with Outliers," Medium, Sep. 05, 2020. https://towardsdatascience.com/limitations-of-arima-dealing-with-outliers-30cc0c6ddf33 (accessed May 12, 2022)

[3] "Energy Charts". $https://www.energy-charts.de/price_{a}vg_{d}e.html$ (accessed May 12, 2022)

Journal articles

[3] Contreras, J., Espinola, R., Nogales, F. & Conejo, A. ARIMA models to predict next-day electricity prices. *IEEE Transactions On Power Systems*. **18**, 1014-1020 (2003

[4] Karabiber, O. & Xydis, G. Electricity Price Forecasting in the Danish Day-Ahead Market Using the TBATS, ANN and ARIMA Methods. *Energies*. **12** (2019), https://www.mdpi.com/1996-1073/12/5/928

[5] Pallonetto, F., Jin, C. & Mangina, E. Forecast electricity demand in commercial building with machine learning models to enable demand response programs. *Energy And AI*. **7** pp. 100121 (2022), https://doi.org/10.1016/j.egyai.2021.100121

[6] Cruz, A., Munoz, A., Zamora, J. & Espinola, R. The effect of wind generation and weekday on Spanish electricity spot price forecasting. *Electric Power Systems Research*. **81**, 1924-1935 (2011)

[7] Pallonetto, F., Galvani, M., Torti, A. & Vantini, S. A Framework for Analysis and Expansion of Public Charging Infrastructure under Fast Penetration of Electric Vehicles. *World Electric Vehicle Journal*. **11** (2020),

https://www.mdpi.com/2032-6653/11/1/18
[8] Chodakowska, E., Nazarko, J. & Nazarko, Ł. Arima models in electrical load forecasting and their robustness to noise. *Energies*.

Conference Paper

[9] Boemer, J., Burges, K., Zolotarev, P., Lehner, J., Wajant, P., Fürst, M., Brohm, R. & Kumm, T. Overview of german grid issues and retrofit of photovoltaic power plants in germany for the prevention of frequency stability problems in abnormal system conditions of the ENTSO-E region continental europe. *1st International Workshop On Integration Of Solar Power Into Power Systems*. **24** (2011)
[10] KumarMahto, A., Biswas, R. & Alam, M. Short term forecasting of agriculture commodity price by using ARIMA: based on Indian market. *International Conference On Advances In Computing And Data Sciences*. pp. 452-461 (2019)
[11] Fragkioudaki, A., Marinakis, A. & Cherkaoui, R. Forecasting price spikes in European day-ahead electricity markets using decision trees. *2015 12th International Conference On The European Energy Market (EEM)*. pp. 1-5 (2015)
[12] Jakaša, T., Andročec, I. & Sprčić, P. Electricity price forecasting — ARIMA model approach. *2011 8th International Conference On The European Energy Market (EEM)*. pp. 222-225 (2011)
[13] Beigaite, R., Krilavičius, T. & Man, K. Electricity price forecasting for nord pool data. *2018 International Conference On Platform Technology And Service (PlatCon)*. pp. 1-6 (2018)
[14] Girish, G., Tiwari, A. & Others A comparison of different univariate forecasting models forSpot Electricity Price in India. *Economics Bulletin*. **36**, 1039

Book, book chapter and manual

[15] GIMON, E. & FELLOW, S. LESSONS FROM THE TEXAS BIG FREEZE. (2021)
[16] Hyndman, R. & Athanasopoulos, G. Forecasting: principles and practice. (OTexts,2018)
[17] Jain, G. & Mallick, B. A study of time series models ARIMA and ETS. *Available At SSRN 2898968*. (2017)