

Some Issues About Iberian Energy Prices

M. Filomena Teodoro[1,2(✉)] and Marina A. P. Andrade[3]

[1] CINAV, Center of Naval Research, Naval Academy, Portuguese Navy, 2810-001 Almada, Portugal
[2] CEMAT, Center for Computational and Stochastic Mathematics, Instituto Superior Técnico, Lisbon University, 1048-001 Lisboa, Portugal
maria.alves.teodoro@marinha.pt
[3] ISTAR, ISCTE, Instituto Universitário de Lisboa, 1649-026 Lisboa, Portugal
marina.andrade@iscte.pt

Abstract. The work described in this article results from a problem proposed by the company EDP - Energy Solutions Operator, in the framework of ESGI 119*th*, European Study Group with Industry, during July 2016. Markets for electricity have two characteristics: the energy is mainly not storable and volatile prices at exchanges are issues to take into consideration. These two features, between others, contribute significantly to the risk of a planning process. The aim of the problem is the short term forecast of hourly energy prices. In present work, ARIMA modeling is considered to obtain a predictive model. The results show that in the time series traditional framework the season of the year, month or winter/summer period revealed significant explanatory variables in the different estimated models. The in-sample forecast is promising, conducting to adequate measures of performance.

Keywords: Electricity Price Forecasting · ARIMA models · MIBEL

1 Introduction

The study of the behavior of time series data is considered one of the current challenges in data mining [2,10]. A wide number of methods for Electricity Price Forecasting (EPF) can be found in literature, for example, in [9] we have a good description of such methods which are labeled in five categories. Data Mining refers the extraction of knowledge by analyzing the data from different perspectives and accumulates them to form useful information which could help the decision makers to take appropriate decisions. Due to the unique behavior of time series data, we can find several references about time series data mining in various application domains, namely survey articles [4,8], PhD thesis where detailed techniques and case studies are analyzed [11], private communications [3], which takes into consideration how Artificial Neural Networks may assist in formulating a GLM, chapters in books [5] where the authors define the major tasks considered by the time series data mining community. Besides that, the

© Springer Nature Switzerland AG 2019
S. Misra et al. (Eds.): ICCSA 2019, LNCS 11621, pp. 122–133, 2019.
https://doi.org/10.1007/978-3-030-24302-9_10

existing research is still considered not enough. In [10] time series data is considered one of the 10 challenges in data mining. In particular the discovery of an interesting pattern, also called motif discovery, is a non-trivial task which has become one of the most important data mining tasks. In fact motif discovery can be applied to many domains [2].

Under the 109th European Study Group with Industry, EDP - Energias de Portugal submitted the mathematical challenge simulating electricity prices not only for risk measures purposes but also for scenario analysis in terms of pricing and strategy. EDP Group is an Energy Solutions Operator which operates in the business areas of generation, supply and distribution of electricity and supply and distribution of gas. EDP with nearly 14 000 MW (2012 update and excluding wind power) of installed capacity in the Iberian Electricity Market – MIBEL[1], is the only company in the Iberian Peninsula with generation, distribution and supply (both electricity and gas) activities in Portugal and Spain.

The objective of the present work is the short term forecast of hourly energy prices. Electricity Price Forecasting (EPF) its a difficult purpose. A wide number of methods have been proposed to EFP. In [9] is described an almost complete review about the enormous quantity of available methods, analyzing their strengths and weaknesses. The author proposes the classification of such methods in four categories: multi-agent models, fundamental models, reduced-form models, statistical models and computational intelligence models.

In present work, usual techniques in time series modelling are used. The ARMA, ARIMA, ARIMAX, SARIMAX are considered useful techniques to obtain a predictive model where its predictive power is discussed.

The outline of this article is developed in four sections. Section 2 makes a brief summary about the time series modeling approach. The Sect. 3 are given more details on the challenge proposed by EDP and on the data provided. Will be presented a summary about exploratory analysis of the data sets provided by EDP and continues with the study on the co-variables that may predict the hourly prices pattern. Also are presented the results of our ARIMA approach. Finally in Sect. 4 conclusions are drawn and suggestions for future work are pointed.

2 ARIMA Approach

The identification of an ARIMA model to model the data can be considered one of the most critical phases when using ARIMA approach. For a stationary time series[2] the selection of the model to be used is mainly based on the estimated auto-correlations and partial auto-correlations, which we will use to compare with the theoretical quantities and identify a possible model for the data.

The auto-regressive models with order p AR(p), the moving average with order q MA(q) and their combination, ARMA(p, q) models have their

[1] http://www.mibel.com/.

[2] A time series is classified as stationary when it is developed in time around a constant mean.

auto-correlation functions (ACF) with a certain specific feature, similarly to a finger print:

The ACF of an autoregressive process with order p is infinite in extent that decays according to a damped exponential/sinusoidal;
The ACF of a moving average with order q process is finite, i.e. presents a cut after the lag q;
The ACP of ARMA process (p, q) is like a mixture of the processes described in previous items, the ACF has infinite that decays according to exponents/damped sinusoidals after the lag $q - p$.

The idea is to identify a pattern that behaves with the same profile that some theoretical model. In particular, the it is useful to identify MA models but it is not so simple to identify other kind of models. As a possible solution, we can compute the partial auto-correlation function (PACF). This function corresponds to the correlation of X_t, X_{t-k+1} removing the effect of the observations $X_{t-1}, X_{t-2}, \ldots, X_{t-k-1}$ and is denoted by ϕ_{kk}. In the case a stationary time series we can use the Yule-Walker equations to compute the PACF. Again, the PACF have a specific profile for each process like a proper finger print:

The PACF of a MA(d) is infinite in extent that decays according to a damped exponential/sinusoidal (similarly to the behaviour of an ACP from a AR(d) process;
The PACF of AR(p) the process is finite, i.e. presents a cut after the lag q, like the behaviour of an ACP from a MA(p) process;
The PACF of ARMA process (p, q) is similar to an ACF from a MA(q) process.

A general method for finding a.c. for a stationary process with f.a. is using the Yule-Walker equations.

This method seems to fail in the case of non-stationary time series (the irregular component is significant). To solve this issue we differentiate the non stationary series so many times (d times) as necessary to get a stationary series. After these differences of successive terms of the chain, applied d times, we can apply the same technique: identify which model(s) are identified from ACF and PACF. A model that represents a nonstationary series, differenciated d times, with an auto-regressive component with order p and a moving average component with order q is representes as an ARIMA (p, d, q).

To estimate the best models between several proposals, we usually apply the information criteria AIC, BIC: the best models have the lowest values of AIC and BIC. Also the log of likelihood function is a good statistic to evaluate the quality of the estimated models: the lowest value means a better model.

After selection, the models need to be validated. One of the rules is to analyze the residuals (the ith residual is the difference between the ith observation and its estimate). Residuals are supposed to be Gaussian and non-correlated. To verify this can be use several testes and other techniques (Llung-Box test, box-Pierce test, Kolmogorov-Smirnov test, Bera and Jarcke test, some graphics, e.g. boxplots, qq plots.

The estimates precision evaluation is another step to include in all process. For that we can compute the usual measures: MAPE, MADE, etc.

In general the procedure of all process is composed by the following iterates:

1. Models formulation: use of ACF and PACF;
2. Models adjustment: estimation of model parameters, application of suitability measures of estimates;
3. Validation of models: selection of variables, diagnostics, residual analysis and interpretation.
4. Analysis of precision and updating the models.

3 Empirical Application

3.1 Exploratory Analysis

Taking into consideration the challenge proposed by EDP, the available data consists in the daily market electricity prices as a strip of prices (one for each hour of the day), all simultaneously observed once at a given time of each day:

$$Y_t = [y_{1t}, y_{2t}, \ldots, y_{nt}], \quad n = 1, \ldots, 24 \text{ (or 23 or 25)}, \quad t = 1, 2, \ldots.$$

In the present work we consider the disaggregated data, i.e., hourly prices and average day price, from January 2008 to June 2016, in a total 3102 observations of the 24 (23 or 25) hours of the day.

In a preliminary exploratory analysis, the data originally provided consisted in a transformed ratio (in what follows named rescaled data) and revealed serious problems which can be visualized in the boxplot diagrams (Fig. 1). The rescaled data has different distributions and a great number of anomalies per hour. These details are also confirmed in Table 1 where some descriptive statistics and tests are summarized. The data detailed description follows what was presented in [7], since we have used the same database set.

From Table 1, we can see the different patterns of dispersion (observe the standard deviation and inter-quartile range columns respectively). Also we confirm that the data does not have normal distribution when we check the Kolmogorov-Smirnov and Jarcke and Bera normality tests.

Consequently, we consider a new data set with the real data. In a preliminary analysis, we have taken the period from 1st January 2008 to 31st December 2010, to exemplify some details and issues and to estimate the initial models considering several covariates of interest.

Since we have a huge dimensional data set, to compare graphically the rescaled data set and the real data set we restrict to the year 2010 the graphics in Fig. 2. We can conclude that rescaled data present a huge quantity of "uncommon" observations each hour of the day with exception of hours 4, 5 and 6. The rescaled data also presents different patterns of dispersion. By other hand, the real data displays unusual observations but in a fewer quantity than in rescaled data. The dispersion of real data presents more homogeneous patterns each hour.

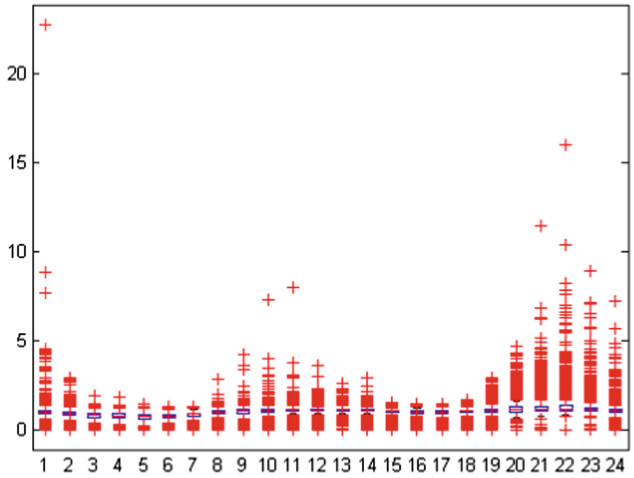

Fig. 1. Boxplot diagrams (rescaled data 01.01.2008–31.12.2016).

Table 1. Descriptive summary (rescaled data 01.01.2008–31.12.2016). Left: Mean, trimmean, media, standard deviation, inter-quartile range. Right: Skewness, kurtosis, Kolmogorov-Smirnov, and Jarcke and Bera normality tests.

Hora	mean	trimmean	median	std	iqr
1.0000	1.0212	0.9908	0.9900	0.5327	0.1800
2.0000	0.8695	0.8872	0.8900	0.2487	0.2000
3.0000	0.7531	0.7943	0.8100	0.2483	0.2100
4.0000	0.7114	0.7541	0.7800	0.2523	0.2300
5.0000	0.6802	0.7230	0.7500	0.2504	0.2400
6.0000	0.7107	0.7573	0.7800	0.2369	0.1900
7.0000	0.8111	0.8594	0.8700	0.2211	0.1600
8.0000	0.9488	0.9773	0.9900	0.2067	0.1600
9.0000	0.9911	1.0163	1.0300	0.2457	0.1900
10.0000	1.0582	1.0666	1.0700	0.2596	0.1500
11.0000	1.0975	1.0988	1.1000	0.2322	0.1200
12.0000	1.0823	1.0896	1.0900	0.1724	0.1100
13.0000	1.0955	1.0998	1.1000	0.1633	0.1200
14.0000	1.0709	1.0807	1.0800	0.1597	0.1100
15.0000	1.0096	1.0282	1.0300	0.1575	0.1000
16.0000	0.9690	0.9973	1.0000	0.1774	0.1300
17.0000	0.9547	0.9872	1.0000	0.1913	0.1500
18.0000	0.9987	1.0209	1.0300	0.1843	0.1400
19.0000	1.0861	1.0715	1.0600	0.2388	0.1700
20.0000	1.1944	1.1275	1.1000	0.3818	0.2500
21.0000	1.2651	1.1717	1.1500	0.4885	0.2575
22.0000	1.3302	1.2027	1.1700	0.6446	0.2400
23.0000	1.2139	1.1298	1.1100	0.4657	0.2000
24.0000	1.0760	1.0336	1.0200	0.3265	0.1700

Hora	skewness	kurtosis	P-value (KS)	P-value (JB)
1.0000	24.5154	927.2962	0	0.0010
2.0000	-0.0747	12.7788	0	0.0010
3.0000	-1.4003	5.3185	0	0.0010
4.0000	-1.2967	4.6556	0	0.0010
5.0000	-1.2724	4.2777	0	0.0010
6.0000	-1.6251	5.3662	0	0.0010
7.0000	-2.2851	8.3236	0	0.0010
8.0000	-1.8786	12.7847	0	0.0010
9.0000	0.0254	24.4068	0	0.0010
10.0000	4.8500	122.5760	0	0.0010
11.0000	8.7122	269.0530	0	0.0010
12.0000	-0.2557	39.8107	0	0.0010
13.0000	-1.5259	24.3516	0	0.0010
14.0000	-1.8895	28.5563	0	0.0010
15.0000	-3.4210	21.3683	0	0.0010
16.0000	-2.9380	15.0985	0	0.0010
17.0000	-2.6087	12.1901	0	0.0010
18.0000	-2.4575	13.7446	0	0.0010
19.0000	0.8480	13.3137	0	0.0010
20.0000	3.1875	20.0743	0	0.0010
21.0000	6.6494	88.1468	0	0.0010
22.0000	8.4558	126.6330	0	0.0010
23.0000	7.0455	75.4992	0	0.0010
24.0000	6.5014	82.1613	0	0.0010

It is possible to find different patterns per day and per hour (see Fig. 3), where, for example, we can see that 22 groups (hours) have mean ranks significantly different from group 1 (hour 1).

Electricity prices can be influenced by the present and past values of various exogenous factors, such as generation capacity, load profiles and meteorological conditions [9], in a preliminary stage we have selected defined and code the following candidates to co-variables: Day of the week – $C_1 = 0, 1, 2, 3, 4, 5, 6$ (Mon, . . . , Sunday); Weekday/Saturday/Sunday – $C_2 = 0, 1, 2$;

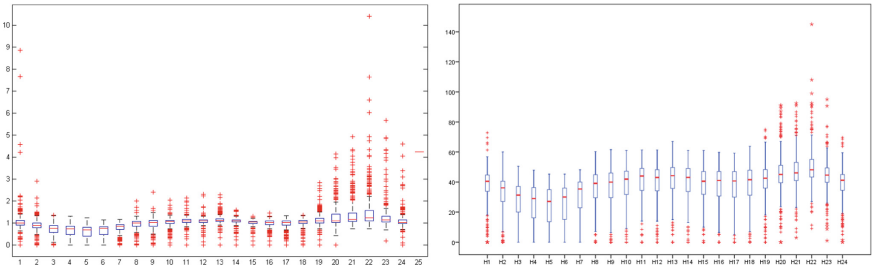

Fig. 2. Boxplot diagrams of rescaled (*left*) and real data (*right*). Time interval: 01.01.2010–31.01.2010.

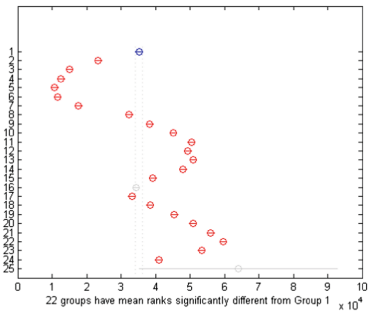

Fig. 3. Real data (01.01.2008–31.01.2008). Mean price per hour.

Weekday/Weekend – $C_3 = 0, 1$; Regular day/holiday – $C_4 = 0, 1$; Season – $C_5 = 0, 1, 2, 3$ (Winter, Spring, Summer, Autumn); Month – $C_6 = 0, \ldots, 11$ (Jan, ..., Dec); Summer/Winter Hour – $C_7 = 0, 1$.

3.2 The Model

Initially, to estimate the model as described before, we considered the time interval from 01/01/2008 to 31/12/2010. The first approach using IBM SPSS Statistics (version 22) was performed with difficulty due the high dimensionality of data. A question that arose was: "Can we reduce the number of components of Y_t?", e.g., are there significant differences between Y_i and Y_j, for $i \neq j$? To solve partially such issue, we try to reduce the 24 h of a day to fewer reference hours. First of all, an analysis of data plot per hour was performed. The graphical representation of data (see Fig. 4) shows similar behavior in some distinct. Identified such similar hours we merge them into an unique interval of similarity. In this way the dimension of data can be reduced, by taking the mean or median or other measure of response variable.

We have selected and defined some time intervals which conduced to the best model performance. In this way, it was reduced the dimension defining the following time intervals: aurora, lunch time and dinner time. Aurora corresponds

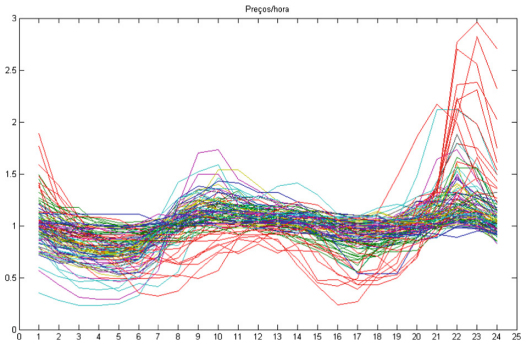

Fig. 4. Data representation (time interval from 01/01/2008 to 31/12/2010).

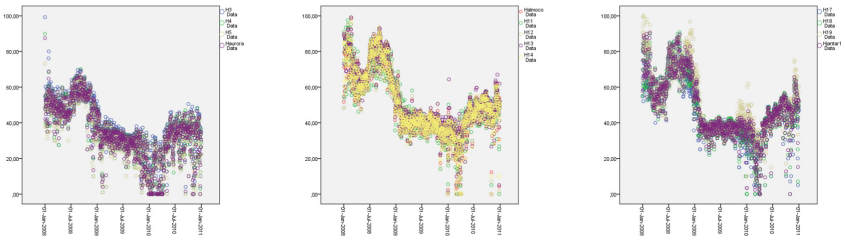

Fig. 5. Overlapped data: Aurora time (left), lunch time (center) and dinner time (right). Time interval from 01/01/2008 to 31/12/2010.

to the hours 3, 4 and 5 respectively. Lunch time merges the hours 11, 12, 13 and 14. Dinner time takes into account hours 17, 18 and 19. When the data is graphically overlapped for each hour in the defined time intervals (see Fig. 5) no significant differences were found.

We studied some possible explanatory variables which can contribute to explain the energy price per hour. In a preliminary stage of the study, using the initial explanatory variables proposed in Subsect. 3.1, an analysis of variance with second order interaction was performed. The best candidates to explanatory variables to include in a model were chosen: C_1, C_4, C_5, C_6, C_7.

It was also considered the fare defined by EDP as possible explanatory variable but it was not significant.

In GLM approach the significant explanatory variables were C_4, C_6, C_7, H_2, H_7, H_8 or H_{16}, H_{20}, H_{22}, H_{23}, H_{24} and lunch time or C_4, C_6, C_7, H_2, H_7, H_8, H_{16}, H_{20}, H_{22}, H_{23}, H_{24} in the several chosen models. In present work, in time series approach, the significant explanatory variables (an some lags) were H_2, H_7, H_8, H_{16}, H_{20}, H_{22}, H_{24}. The selected model has similar explanatory variables to some models obtained by GLM in [7].

Considering the obtained results as indicators, we can conclude that some of the explanatory variables proposed initially were not relevant for dependent variable, such as, EDP fares, Portuguese holidays (maybe the Iberian holidays can have some relevance, and not just the Portuguese ones). Also, some periods of time can be drop off as relevant explanatory variables, such as dinner time or some others. The season, month or winter/summer time period revealed significant explanatory variables in the different estimated models.

Using this preliminary model estimation as starting point, we repeated all estimation process considering a more recent sample so we could compare with the results published in [6,7]. The ARIMA model was estimated using hourly prices from 10/03/2014 to 29/5/2016. The remaining sample, from 30/05/2014 to 28/06/2016, was used to evaluate the forecasting performance of the selected model. To asses the in-sample prediction quality of the model, we use the Mean Absolute Percentage Error (MAPE) and the Root Mean Square Error (RMSE). The mean values of MAPE and RMSE are with accordance withe the results published in [6,7].

We have selected as preliminaries explanatory variables the same used earlier also considered in [6,7], where its done a VAR approach and a GLM approach respectively. There were estimated of model parameters and analyzed the suitability measures of estimates. The selection and validation of models such as selection of variables, diagnostics, residual analysis and interpretation was concluded. All models obtained good significant results. Here we present the model with similar performance (estimation and forecasting) and with best R^2.

In Fig. 6 we can find the details of the model selected by time series approach.

ARIMA Model Parameters

					Estimate	SE	t	Sig.
DIF_1_Average-Model_1	DIF_1_Average	No Transformation	AR	Lag 7	.102	.037	2.767	.006
			MA	Lag 1	.888	.017	51.352	.000
	H2	No Transformation	Numerator	Lag 0	.156	.007	22.365	.000
				Lag 1	.156	.007	23.078	.000
	H7	No Transformation	Numerator	Lag 0	.127	.011	11.773	.000
				Lag 1	.129	.011	12.175	.000
	H8	No Transformation	Numerator	Lag 0	.147	.008	17.598	.000
				Lag 2	.141	.008	16.703	.000
			Denominator	Lag 1	-.928	.023	-40.121	.000
	H16	No Transformation	Numerator	Lag 0	.262	.006	43.271	.000
				Lag 1	.263	.006	44.348	.000
	H20	No Transformation	Numerator	Lag 0	.111	.007	16.844	.000
				Lag 1	.112	.007	17.142	.000
	H22	No Transformation	Numerator	Lag 0	.086	.008	10.825	.000
				Lag 2	.090	.008	11.538	.000
			Denominator	Lag 1	-.654	.057	-11.393	.000
				Lag 2	.208	.041	5.095	.000
	H24	No Transformation	Numerator	Lag 0	.056	.007	8.495	.000
				Lag 1	.052	.007	7.856	.000

Fig. 6. Estimated ARIMAX model.

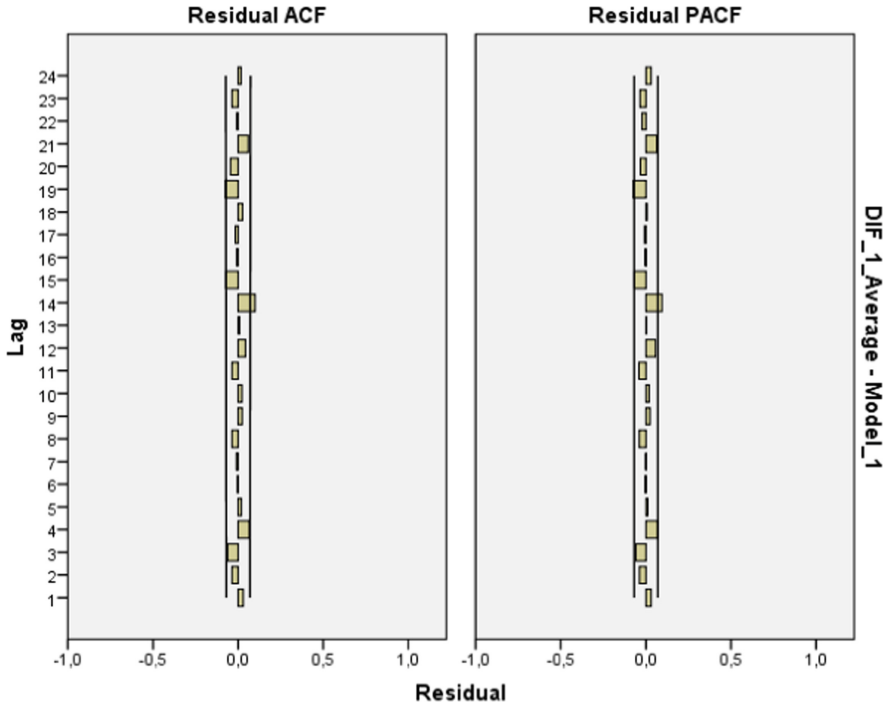

Fig. 7. Estimated ACP and PACF.

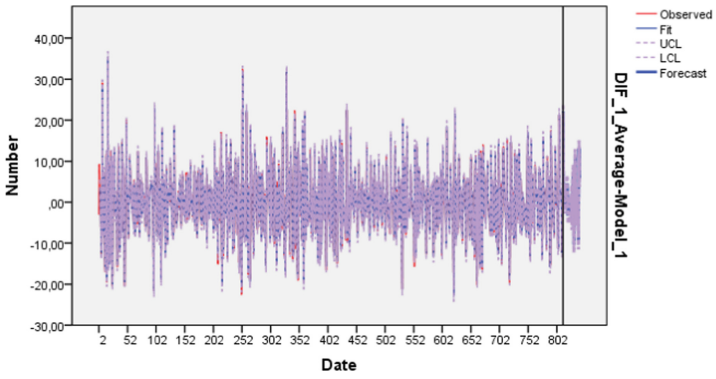

Fig. 8. Residuals representation. Estimation period: 10/03/2014 to 29/5/2016.

The diagnostic analysis and selection of the order of the models was done but we do not reproduce with detail such work. The ACF and PACF are displayed in Fig. 7. We can see that these functions are in agreement with the estimated model (ARIMA(7, 1, 1). It seems in disagreement with Fig. 6 where is displayed a model ARIMA(7, 0, 1), but the considered time series was differentiated previously (one time) to the model estimation.

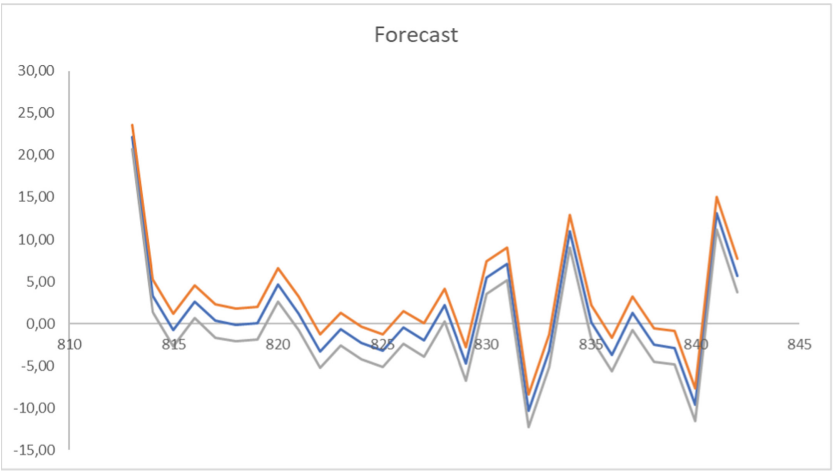

Fig. 9. Forecast (blue), UPL (orange), LPL (grey). Period: 30/05/2016 to 28/06/2016. (Color figure online)

Table 2. Forecast errors - MADE and RMSE. Period: 30/05/2016 to 28/06/2016.

Date	MADE	RMSE	Date	MADE	RMSE
30/05/2016	2,29	2,285	**14/06/2016**	5,36	5,362
31/05/2016	0,18	0,184	**15/06/2016**	4,74	4,735
01/06/2016	2,44	2,445	**16/06/2016**	7,72	7,720
02/06/2016	0,70	0,700	**17/06/2016**	8,72	8,718
03/06/2016	0,84	0,841	**18/06/2016**	5,15	5,154
04/06/2016	0,02	0,024	**19/06/2016**	11,05	11,047
05/06/2016	4,62	4,620	**20/06/2016**	0,72	0,722
06/06/2016	1,53	1,532	**21/06/2016**	3,50	3,499
07/06/2016	3,37	3,370	**22/06/2016**	0,34	0,335
08/06/2016	1,07	1,067	**23/06/2016**	0,96	0,958
09/06/2016	1,43	1,434	**24/06/2016**	2,98	2,980
10/06/2016	1,32	1,321	**25/06/2016**	10,03	10,035
11/06/2016	1,94	1,936	**26/06/2016**	12,83	12,828
12/06/2016	0,82	0,816	**27/06/2016**	5,50	5,496
				2,366	**2,365**

When we analyze the graphics in Fig. 8, we can conclude that the residuals behavior is in accordance with the expected (null mean and constant variance). Also their Gaussian distribution was not rejected statistically. In Fig. 9 it is presented the forecast from 30/05/2016 to 28/06/2016. The model evidences adequate performance estimation. We almost cant distinguish the original series and the predicted series.

From Table 2 we can analyze the quality of prediction in-sample using the MAPE and RMSE. We can conclude that the forecasting quality is promising. Notice that the RMSE values are in accordance with the results obtained using the VAR and the GLM approach [6,7].

4 Conclusions and Recommendations

The challenge proposed by EDP consisted in simulating electricity prices not only for risk measures purposes but also for scenario analysis in terms of pricing and strategy. Data concerning hourly electricity prices from 2008 to 2016 were provided by EDP.

The data were explored using different statistical software, namely IBM SPSS Statistics and Matlab. In this work a time series approach was considered. The season of the year, month or winter/summer period revealed not significant explanatory variables in the different estimated models in opposite to GLM approach [7]. Also, We have selected a model that not includes the reduced form of day hours (aurora time, lunch time, dinner time). From Table 2 we can analyze the quality of prediction in-sample by MADE and RMSE. We can conclude that the forecasting quality is promising. When compared with multivariate approach using the VAR approach and GLM approach [6,7] for the same period (from 30/05/2016 to 28/06/2016) the RMSE values are in accordance with the RMSE computed using the VAR method. Although the forecast do not exactly replicate the real price the results are quite promising. The introduction of other co-variables, such as oil price, gas price, wind energy production, other meteorological variables, would certainly improve the model and the forecast. Univariate time series approach still needs to be improved in the sense of comparison with other similar models. Others methods should be explored. It is still possible to explore another models in future work.

EPF literature has mainly concerned on models that use information at daily level, however this particularly problem proposed is interested in forecasting intra-day prices using hourly data (disaggregated data), maybe it is necessary to consider models that explore the complex dependence structure of the multivariate price series. The problem of modeling distributional properties of energy prices can be classified in three main classes: reduced form models, forward price models and hybrid price models [1]. Temporal Distribution Extrapolation is another possible idea for our future work.

Acknowledgements. This work was supported by Portuguese funds through the *Center of Naval Research* (CINAV), Portuguese Naval Academy, Portugal and *The Portuguese Foundation for Science and Technology* (FCT), through the *Center for Computational and Stochastic Mathematics* (CEMAT), University of Lisbon, Portugal, project UID/Multi/04621/2019.

References

1. Eydeland, A., Wolyniec, K.: Energy and Power Risk Management: New Developments in Modeling, Pricing, and Hedging. Wiley, Hoboken (2003)
2. Fu, T.: A review on time series data mining. Eng. Appl. Artif. Intell. **24**(1), 164–181 (2011)
3. Mulquiney, P.: Combining GLM and data-mining techniques for modelling accident compensation data (private communication)
4. Pal, S.H., Palet, J.N.: Time series data mining: a review. Bin. J. Data Min. Netw. **5**, 01–04 (2015)
5. Ratanamahatana, C.A., Lin, J., Gunopulos, D., Keogh, E., Vlachos, M., Das, G.: Mining time series data. In: Maimon, O., Rokach, L. (eds.) Data Mining and Knowledge Discovery Handbook, pp. 1049–1077. Springer, Boston (2010). https://doi.org/10.1007/978-0-387-09823-4_56
6. Costa e Silva, E., Borges, A., Teodoro, M.F., Andrade, M.A.P., Covas, R.: Time series data mining for energy prices forecasting: an application to real data. In: Madureira, A.M., Abraham, A., Gamboa, D., Novais, P. (eds.) ISDA 2016. AISC, vol. 557, pp. 649–658. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-53480-0_64
7. Teodoro, M.F., Andrade, M.A.P., Silva, E.C., Borges, A., Covas, R.: Energy prices forecasting using GLM. In: Oliveira, T.A., Kitsos, C.P., Oliveira, A., Grilo, L. (eds.) Recent Studies on Risk Analysis and Statistical Modeling. CS, pp. 321–334. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-76605-8_23
8. Vasimalla, K.: A survey on time series data mining. Int. J. Innov. Res. Comput. Commun. Eng. **2**(5), 170–179 (2014)
9. Weron, R.: Electricity price forecasting: a review of the state-of-the-art with a look into the future. Int. J. Forecast. **30**(4), 1030–1081 (2014)
10. Yang, Q., Wu, X.: 10 challenging problems in data mining research. Int. J. Inf. Technol. Decis. Mak. **5**(04), 597–604 (2006)
11. Zhu, Y.: High performance data mining in time series: techniques and case studies. Ph.D. thesis, Department of Computer Science, University of New York (2004)