A Longitudinal Model for MIBEL Energy Prices

ANA BORGES
ESTG, Polytechnic of Porto
CIICESI
4610-156 Felgueiras
PORTUGAL
aib@estg.ipp.pt

ELIANA COSTA E SILVA
ESTG, Polytechnic of Porto
CIICESI
4610-156 Felgueiras
PORTUGAL
eos@estg.ipp.pt

RICARDO COVAS
EDP - Energias de Portugal
CMA-UNL
1249-300 Lisboa
PORTUGAL
Ricardo.Covas@edp.pt

Abstract: We propose to contribute to the problematic of Electricity Price Forecasting with a longitudinal statistical approach. We focus our interest on forecasting intra-day prices using hourly data (disaggregated data) in a multivariate approach rather than in the usually used univariate approach, by adjusting a mixed-effects longitudinal model to the Iberian Electricity Market hourly prices from January 1th 2015 to June 26th 2016, in a total of 13 032 observations. Results indicate that a longitudinal approach considering a mixed-effects model, with month and weekday as fixed effects, hour group as random effect and an AutoRegressive component of order 7 describing the within hour dependence, yield a model that explains the intra-day and intra-hour dynamics for the electricity hourly prices.

Key–Words: Electricity Price Forecasting, Longitudinal Mixed-effects Model, MIBEL.

1 Introduction

Under the 109th European Study Group with Industry (ESGI 119)[1], the company EDP - Energias de Portugal submitted the mathematical challenge of 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[2] of installed capacity in the Iberian Electricity Market – MIBEL[3], is the only company in the Iberian Peninsula with generation, distribution and supply (both electricity and gas) activities in Portugal and Spain.

The search for accurate models to predict price movements can help develop profit-maximizing trading strategies and optimal bidding techniques [5]. In fact, important activities such as bidding strategies rely, nowadays, on price forecast information to improve decision making. There is, in fact, an increased significant relevance of short term electricity forecasting in the energy price research for trading and bidding, since most of the electricity markets across the world are liberalized [5].

The present work proposes a contribution to the stated challenge, by presenting a longitudinal statistical approach to Electricity Price Forecasting (EPF).

Statistical EPF models are mainly inspired from economics literature (such as game theory models and time-series econometric models). Murthy et. al., in [16], summarize a selection of finance and econometrics inspired literature on spot electricity price forecasting. As [9] exposes, the statistical EPF models, for short term price forecasting, commonly used are:

- univariate AutoRegressive models (AR);
- AutoRegressive Moving Average models (ARMA);
- AutoRegressive Integrated Moving Average (ARIMA);
- Seasonal ARIMA models (SARIMA).

Reported works have already been developed on electricity price forecasting, applying statistical techniques concerning the Portuguese and Spanish electricity market - the market operated by the group EDP [3].

For example, [2] provides a method to predict next-day electricity prices of mainland Spain market based on the ARIMA methodology.

Also, [14] apply forecasting factor models to the market framework in Spain and Portugal. More recently, [15] proposes an enhanced hybrid approach composed of an innovative combination of wavelet transform, differential evolutionary particle swarm optimization,

[1]http://www.estgf.ipp.pt/esgi/
[2]2012 update and excluding wind power
[3]http://www.mibel.com/

and an adaptive neuro-fuzzy inference system to forecast short-term electricity market prices signals in different electricity markets (and also wind power) in Portugal.

EPF literature has mainly concerned on models that use information at daily level (aggregated data). However, we focus our interest on analysing intra-day prices using hourly data (disaggregated data) in a multivariate approach rather than in the usually used univariate approach. Hence, we consider a longitudinal model that is able to incorporate the complex dependence structure of a multivariate price series.

A day-ahead market consists in a system where agents submit their bids and offers for the delivery of electricity for each hour of the next day before a certain market closing time [9]. Hourly prices for next day delivery are determined at the same time.

The present work studies the electricity prices of MIBEL. The daily market electricity prices can be given as a strip of prices (one for each hour of the day), all simultaneously observed once at a given time of each day. Therefore, the daily market prices can be interpreted as a Longitudinal data.

Considering, in this particular analysis, the hours of the day as subjects, and the electricity prices for those hours for each day throughout the year, we are in the presence of balanced longitudinal data, i.e., repeated measurements for each subject (hour), taking at the same moment (day).

Longitudinal mixed-effects models are extremely popular in social, biological sciences and econometrics (often termed panel data in this last particulary area) and their popularity is explained by the flexibility they offer in modeling the within-group correlation often present in grouped data [4].

One example of their use in this context, although concerning to hydro power, is the work of [7] where they estimate hourly prices through panel data (longitudinal data) methodology, for bidding optimization.

Huisman et. al. [6] give an interest insight and justification on the application of a panel model to describe the dynamics in day-ahead hourly prices. In particular for three European wholesale power markets: the APX (the Netherlands), EEX (Germany) and PPX (France). They clarify that the dynamics of hourly electricity prices does not behave as a time series process. Instead, these prices should be treated as a panel in which the prices of 24 cross-sectional hours vary from day to day. Explaining that, within the biding process, a trader uses exactly the same information to set the price for hour x as it uses to set the price for hour y, where $x \neq y$. Proceeding to the next day, the information set updates, but it updates simultaneously for hour 1 through 24. Hence, they conclude that hourly prices within a day behave cross-sectionally and hourly dynamics over days behave according to time-series properties.

Our present work is an extension of [3] analysis, where a vector autoregressive model with exogenous variables (VARX) was proposed. In order to compare the VARX approach with the longitudinal here proposed, we also analysed the hourly prices from 2015 until 2016.

Since longitudinal models rely on the assumption, among others, of independent subjects (in this particular case, hours), we initially tested for correlation among the time series of electricity prices for the 24 hours of the day, by graphical interpretation of the correlogram (a correlation matrix) and by preforming a factorial analysis.

Results show the existence of the following three independent groups of hours:

A) from the 1st until the 7th hour;

B) from the 8th until the 18th hour;

C) from the 19th until the 24th hour.

We then adjusted a mixed-effects longitudinal model to the data, considering the three groups as subjects and, as reference time, the time since the first day of the year.

By fitting a mixed-effects model to these data we are able to make inferences about the fixed effects, which represent average characteristics of the population represented by these groups of hours, and the variability amongst these three group of hours.

This paper is organized as follows: Section 2 presents a summarized description of the database; Section 3 explains the methodology and Section 3 exposes the main results, ending with the Conclusion Section where suggestions for future work are also pointed.

2 Electricity Database

MIBEL, created in 2004, resulted from the cooperation between the Portuguese and Spanish Governments with the aim of promoting the integration of both countries' electrical systems, involving the integration of their respective electric power systems and their previous electricity markets.

MIBEL allows any consumer in the Iberian region (mainland of Portugal and Spain) to purchase electrical energy under a free competition regime from any producer or retailer acting in that region [13]. As [13] explain, daily and intraday markets are organized in a daily session, where next-day sale and electricity purchase transactions are carried out, and in six intraday sessions that consider energy offer and demand, which may arise in the hours following the daily viability schedule fixed after the daily session.

The daily market electricity prices can be given as a strip of prices (one for each hour of the day), all simultaneously observed once at a given time of each day:

$$Y_t = [y_{1t}, y_{2t}, \ldots, y_{nt}],$$

where

$$n = 1, \ldots, 24 \quad \text{(or 23 or 25)},$$

$$t = 1, 2, \ldots.$$

Therefore, the daily market prices can be interpreted as a multivariate time series.

In order to understand the complexity of the hourly electricity prices, we present in Figure 1, as an illustrative example, the hourly prices for the 31 days of January 2016. This figure suggests a common pattern among the different hours of the day, although each intraday period displays a rather distinct price profile, reflecting the somehow complex dynamics of price over time.

For a deeper analysis of the pattern, Figure 2 presents the 10th up to 90th percentiles MIBEL hourly prices divided by price average of each day. From it we can see the characteristic price profile.

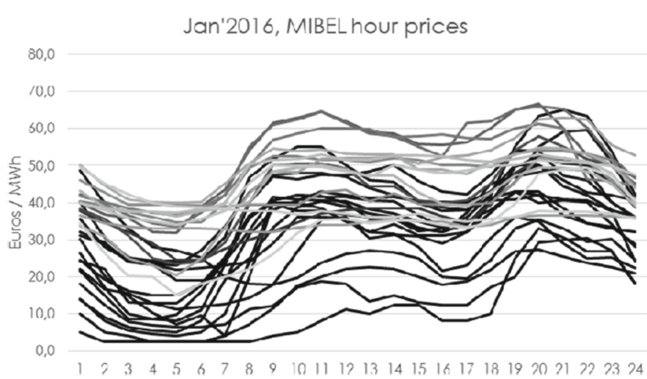

Figure 1: MIBEL prices from January 2016 (graphic provided by EDP during the 119th ESGI).

The data analyzed in this work consists of disaggregated data, i.e., hourly day prices. Our sampling period was specified from January 1th 2015 to June 26th 2016, yielding a total of 13 032 observations.

Given that the present analysis precedes our work with this same data exposed in [3], we have considered the prices on the logarithmic scale as done on

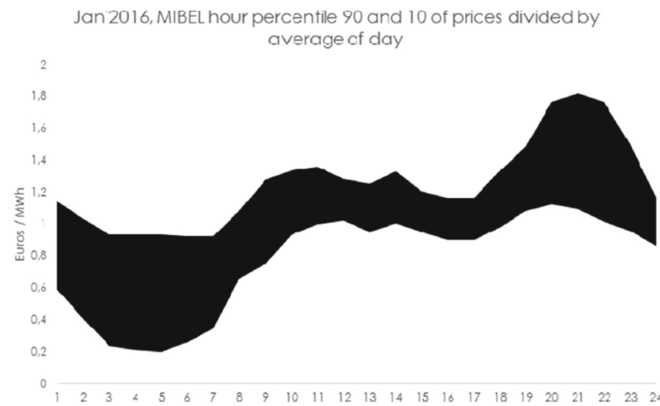

Figure 2: Percentiles 10th to 90th of MIBEL hourly prices divided by price average of each day (graphic provided by EDP during the 119th ESGI).

the referred analysis.

Figure 3 presents the MIBEL daily prices of the 24 hours from 01/01/2015 to 26/06/2016, along with the smooth spline (green color) describing the mean progression in time of daily electricity price. Since the days 29/03/2015 and 27/03/2016 only presented 23 legal hours, these missing values were filled with the previous hour price value, with the assumption that the current data will be similar to the previous ones (see also [3]).

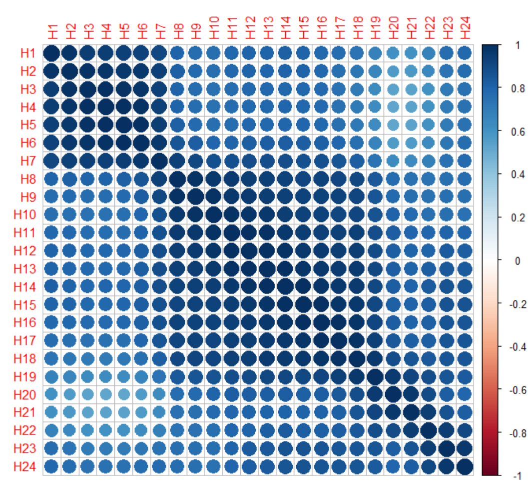

Figure 4: Correlogram.

Since longitudinal models rely on the assumption, among others, of independent subjects (in this particular case, hours), we initially tested for correlation among the time series of electricity prices for

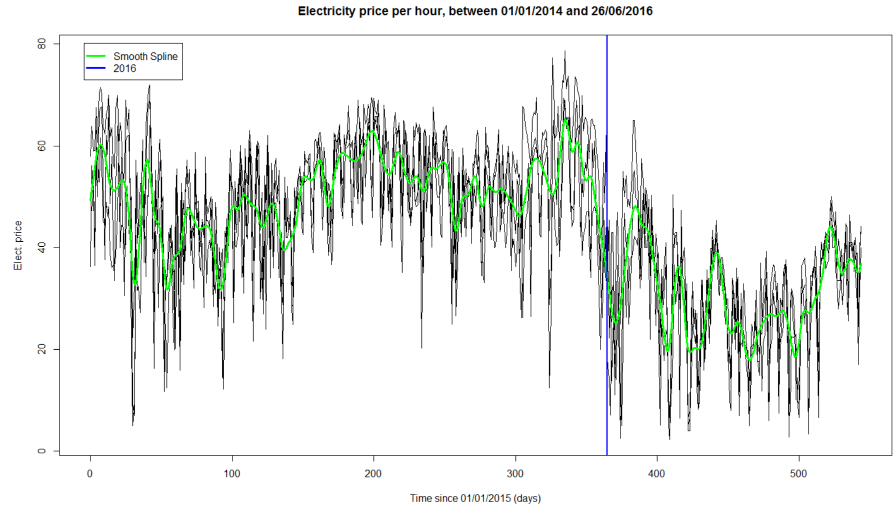

Figure 3: MIBEL hourly prices for each of the 24 hours of the day from 01/01/2015 to 26/06/2016 (data provided by EDP during the 119th ESGI).

the 24 hours of the day. A graphical representation of the correlation between the 24 time series, presented on Figure 4, where larger and darker circles represent higher correlations, shows evidence for a clear cross-sectional correlation structure in hourly electricity prices.
Similar results were found in [6], where it is justified that this effect arises because consumption and capacity flows over the hours. The author explains that if reserve capacity is low in one hour it will probably be low in the next hour as well and if demand is high in one hour it will probably be high in the next hour as well. This is also justified by the complexity of the electricity generation and the impossibility of storing energy.
Taking that result into account, we grouped the 24 hours in the following three main independent sets:

A) from the 1st until the 7th hour;

B) from the 8th until the 18th hour;

C) from the 19th until the 24th hour.

Hence, in the present analysis we consider these three groups, A, B and C as independent subjects. The heterogeneity between the price dynamics in time for each group of our is explicit in Figure 5, that presents three smooth splines describing the mean progression in time of electricity prices throughout the period analysed. The mean progression in time for electricity prices for group A (green color), that includes hours from midnight until 7, presents lower electricity price values, comparing to the other two groups. Being the group C, that represents the last hours of the day, the one that presents higher electricity price values on its progression in time.

3 Methodology

Longitudinal data is usually characterized as response variables that are measured repeatedly through time for a group of individuals.
They are useful since they can provide detailed representations of characteristics that are unique to each subject, thus accounting for a possible problem of heterogeneity. The main characteristic of longitudinal models is that they model both the dependence among the response on the explanatory variables and the autocorrelation among the responses. Ignoring correlation in longitudinal data could lead to incorrect inferences about the regression coefficients, inefficient estimates of the coefficients and, also, sub-optimal protection against biases causes by missing data [4].
Particulary in the present study, we are dealing with hours of the day as subjects, and interested in modelling the progression in time of the electricity prices for those hours for each day throughout the year. We are analysing balanced longitudinal data, i.e, repeated measurements for each subject (hour), taking at the same moment (day).
A mixed-effects model was adjusted to the data corresponding to the hourly prices from 01/01/2015 to 28/06/2016 divided in the three groups mentioned in the previous section. The analysis was performed using R Statistical Software (version 3.3.0)[10], in par-

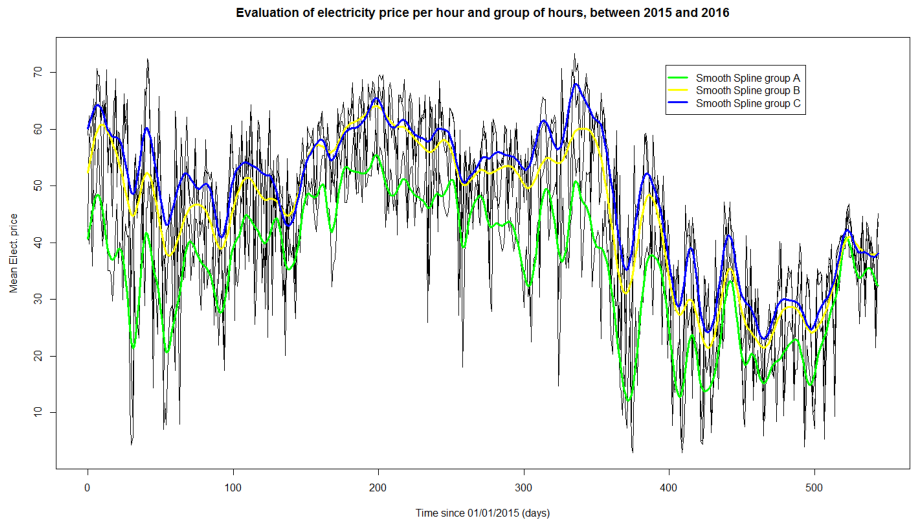

Figure 5: Individual mean progression in time for each group of hour: A, B and C

ticular making use of the *nlme* package [11].
As explained in the previous section, there is, in our perspective, a difference on the dynamic of three distinct group of hours, A, B and C. As so, to account for the variability between the three groups of hours, we adjusted the linear mixed-effects model proposed by Laird and Ware (1982) where the n_i dimensional response vector y_i for the i group is given by:

$$y_i = X_i\beta + Z_i b_i + \epsilon_i, \quad (1)$$
$$i = 1, ..., M,$$

where the p-dimensional vector of random effects b_i are M i.i.d realizations of $N(0, \Sigma)$, the n_i dimensional within-group error vector ϵ_i with a spherical Gaussian distribution with mean zero and variance $\sigma^2 I$.
The p dimensional vector of fixed effects associated with the known fixed-effect covariates matrix X_i (of type $n_i \times p$), and Z_i (of type $n_i \times p$) is the random-effects covariates matrix.
The random effects b_i and the within-group errors ϵ_i are assumed to be independent for different groups and to be independent of each other for the same group, i.e., the group of hours A, B and C are assumed independent as so are the days in each group of hours.
For fixed-effects we considered the weekday and the month among other covariates such as season of the year, a dummy variable comparing weekday with weekend. However, only the first two were significant in the model.
The pattern of the empirical autocorrelation function [1] of the within-group residuals, presented in Figure 6, suggests a strong correlation between two measurements in time of lag 7.
Hence, we extend the basic linear mixed-effects model to take into account a serial correlation among observations in the same group of hours. Hence, to model dependence among the within-group errors we included an AutoRegressive component of order 7 in our model (1).

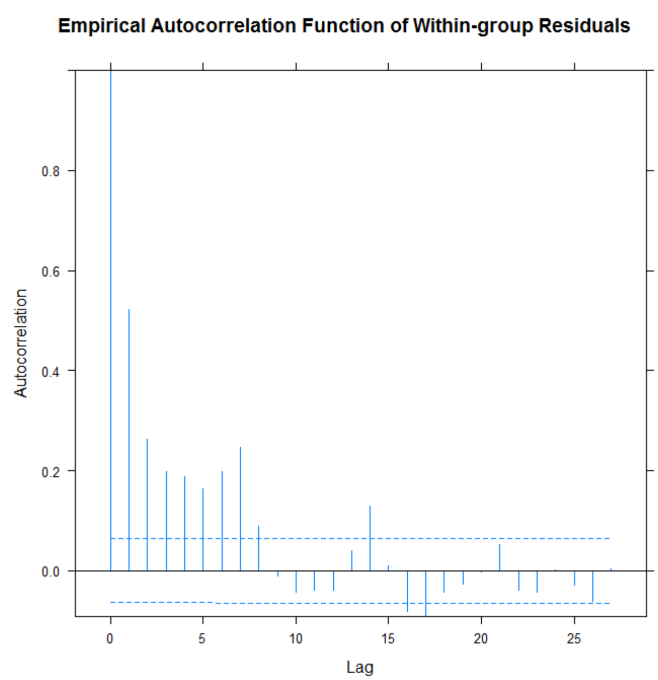

Figure 6: Empirical autocorrelation function

As [8] explains, the general within-group correlation structure, for $i = 1, ..., M$ and $j, j' = 1, ..., n_i$ can be expressed as:

$$cor(\epsilon_{ij}, \epsilon_{ij'}) = h[d(p_{ij}, p_{ij'}), \rho] \tag{2}$$

where ρ is a vector of correlation parameters and $h(.)$ is a correlation function, continuous in ρ, such that for two identical positions vectors $p_{ij} = p_{ij'}$ we have a correlation of $h(0, \rho) = 1$.
The autoregressive model of order p, $AR(p)$ expresses the current observation as a linear function of previous observations as:

$$\epsilon_t = \phi_1 \epsilon_{t-1} + ... + \phi_p \epsilon_{t-p} + a_t \tag{3}$$

where a_t is a homoscedastic noise term, centered at 0 and assumed independent of the previous observations. The coefficients ϕ_i describe the dependence between the price on day t with previous days.
Particulary, concerning the autoregressive model of order 7, AR(7), the correlation function can be defined recursively through the difference equation [1]:

$$h(k, \phi) = \phi_1 h(|k-1|, \phi) + ... + \phi_7 h(|k-7|, \phi), \tag{4}$$

$$k = 1, 2, ...$$

Note that in a previous analysis [3] the autoregressive coefficients of an adjusted VARX(7,0) explained the existence of dependence within the hourly prices, which corroborates our choice on the serial correlation structure.
The estimates of the model were obtained by maximum likelihood methodology.

4 Main Results

Table 1 summarizes and compares the estimated parameters for the two longitudinal models fitted:

- RESC, which incorporates both the random effect intercept and a AR(7) serial correlation within the group of hours;
- RE, which only accounts for a random effect intercept between the group of hours.

Comparing the values of the loglikelihood of both models and the respective Akaike information criterion (AIC), we infer that the more complex model, the one that incorporates a random intercept effect and a AR(7) serial correlation within the group of hours, provides an better fit. Hence, is more adequate to describe the progression in time of MIBEL electricity prices.
The reference (baseline) covariate categories for the fixed effects are Monday, as weekday, and January, as month.

Table 1: Estimated Parameters Values for MIBEL Electricity Prices Longitudinal Models

	RESC Model		**RE Model**	
	Est	**p-value**	**Est**	**p-value**
Intercept	4.438	< 0.001	4.491	< 0.001
Time	-0.0014	< 0.001	-0.0014	< 0.001
Weekday				
Tuesday	0.056	< 0.01	0.056	0.042
(Wednesday)	0.042	0.1203	0.043	0.1253
(Thursday)	0.048	0.0875	0.049	0.0813
Friday	0.059	0.0369	0.058	0.0371
Saturday	-0.112	< 0.001	-0.113	< 0.001
Sunday	-0.328	< 0.001	-0.328	< 0.001
Month				
(March)	-0.023	0.7445	-0.110	< 0.001
(April)	-0.020	0.7715	-0.116	< 0.001
(May)	0.004	0.9493	-0.050	0.1145
June	0.307	< 0.001	0.329	< 0.001
July	0.382	< 0.001	0.378	< 0.001
August	0.428	< 0.001	0.368	< 0.001
September	0.374	< 0.001	0.322	< 0.001
October	0.381	< 0.001	0.327	< 0.001
November	0.413	< 0.001	0.396	< 0.001
December	0.500	< 0.001	0.442	< 0.001
σ^2	0.0205		0.0219	
ϕ_1	0.5040			
ϕ_2	-0.0590			
ϕ_3	0.0419			
ϕ_4	0.0564			
ϕ_5	-0.0159			
ϕ_6	0.0544			
ϕ_7	0.1148			
τ^2	0.0916		0.0893	
Log Likelihood	-72.16		-352.29	
AIC	200.2749		746.5801	

Analysing the categories that have a significant effect on the mean progression in time of the electricity prices, we can infer that weekdays such as Tuesday and Friday have an increasing effect on the progression of electricity price log values (positive values of the coefficient estimates), comparing to the Monday (the reference category). While Saturday and Sunday have a decreasing effect on the electricity price log values (negative values of the coefficient estimates), in comparison to Monday.
In this aspect, the main difference between the two models, RESC and RE, relies on the effect of Tuesday, which ceases to have on the latter.
Furthermore, results suggest that it is not expected

a significant difference between electricity prices between March (p-value = 0.7445), April (p-value = 0.7715) and May (p-value = 0.9493) comparing to January (the reference category). On the other hand, months from June to December will represent an increase on the starting value of the mean progression in time of the electricity prices, since they present a positive value on the coefficient estimates. Nevertheless, it is worth of pointing out that, for the RE model, it resulted that there is a significant difference between electricity prices between March and April comparing to the reference category.
For model diagnose we constructed the boxplot of residuals by group of hour (A, B and C) for the RESC model. This plot is useful for verifying that the errors are centered at zero and have constant variance across the three hour groups, and are independent of the group levels.

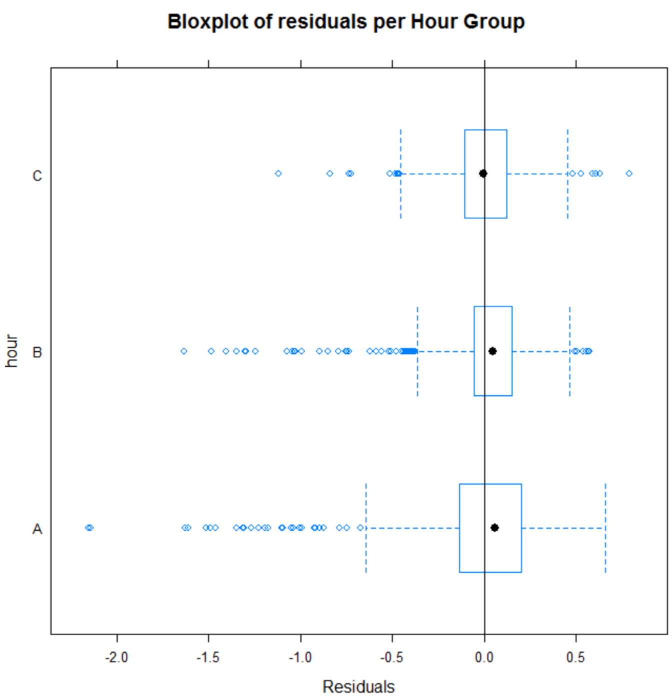

Figure 7: Boxplot of residuals per group of hour for RESC longitudinal model

Analysing the boxplots presented in Figure 7, we can observe that the residuals are centered at zero, but that the variability changes with group. However, the standardized residuals are small, suggesting that the model is successful in explaining the variation of electricity price.
Which can be also seen by inspection of the plot of the observed responses versus the within-group fitted values presented in Figure 8. The fitted values are in close agreement with the observed electricity prices.

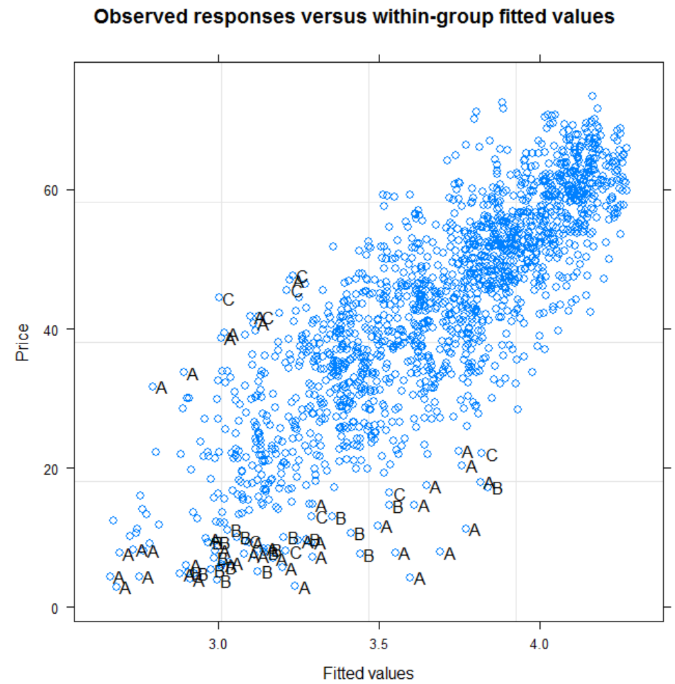

Figure 8: Scatterplot of observed responses versus within-group fitted values for RESC longitudinal model.

5 Conclusions and Future Work

The challenge proposed by EDP at ESGI119 consisted in predicting and modelling the electricity prices variation, not only for risk measures purposes, but also for scenario analysis in terms of pricing and strategy. It is understandable that fitting an adequate model to predict and describe price dinamic can lead to profit-maximizing trading strategies and optimal bidding techniques.

Data provided by EDP, concerning hourly electricity prices from 2015 to 2016, was analysed making use of a longitudinal model.

Results show that the longitudinal modeling approach considering a mixed-effects model, with the season of the year and the type of day (weekday versus weekend day) as fixed effects and the hour group as random effect, yield a model that explains the intra-day and intra-hour dynamics of the hourly prices.

By fitting a mixed-effects model to these data we are able to make inferences about the fixed effects, which represent average characteristics of the population represented by these groups of hours, and the variability amongst the three group of hours.

In subsequent work, we aim to improve the model performance by testing and introducing exogenous covariates such as fuel prices or the capacity surplus. Also, we intend to forecasts future MIBEL electricity prices based on the theory of best linear unbiased predictors, BLUPs, extending the work of [12].

Acknowledgements: This work was partially supported by ESGI 119 an initiative supported by COST Action TD1409, Mathematics for Industry Network (MI-NET), COST is supported by the EU Framework Programme Horizon 2020. E. Costa e Silva and A. Borges were supported by Center for Research and Innovation in Business Sciences and Information Systems (CIICESI), ESTG - P.Porto.

References:

[1] Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1994). Time Series Analysis: Forecasting and Control, 3rd ed., Holden-Day, San Francisco.

[2] Contreras, J., Espinola, R., Nogales, F.J., Conejo, A.J. (2003) ARIMA models to predict next-day electricity prices. IEEE Trans. Power Syst. 18(3), 10141020.

[3] Costa e Silva, E., Borges, A., Teodoro, M.F., Andrade, M.A.P., Covas, R. (2017) Time Series Data Mining for Energy Prices Forecasting: An Application to Real Data. IN A.M. Madureira et al. (Eds.), Intelligent Systems Design and Applications, Advances in Intelligent Systems and Computing 557.

[4] Diggle, P., Heagerty, P., Liang, K.Y., Zeger, S. (2002). Analysis of Longitudinal Data, 2nd edition. Oxford, England: Oxford University Press.

[5] dos Santos Coelho, L., Santos, A. A. (2011). A RBF neural network model with GARCH errors: application to electricity price forecasting. Electric Power Systems Research, 81(1), 74-83.

[6] Huisman, R., Huurman, C., Mahieu, R. (2007). Hourly electricity prices in day-ahead markets. Energy Economics, 29(2), 240-248.

[7] Klaboe, G. (2015). Forecasting hourly electricity prices for bidding optimization. In European Energy Market (EEM), 2015 12th International Conference on the IEEE,1-5.

[8] Pinheiro, J. C., Bates, D. M. (2000). Mixed-effects models in S and S-PLUS. New York: Springer.

[9] Weron, R. (2014) Electricity price forecasting: A review of the state-of-the-art with a look into the future. International Journal of Forecasting, 30(4), 1030-1081.

[10] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016.

[11] Pinheiro J, Bates D, DebRoy S, Sarkar D and R Core Team, (2014). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-118, http://CRAN.R-project.org/package=nlme.

[12] Frees, E.W. and Miller, T.W. (2004). Sales forecasting using longitudinal data models. International Journal of Forecasting 20, 99114.

[13] Monteiro, C., Ramirez-Rosado, I. J., Fernandez-Jimenez, L. A., and Conde, P. (2016). Short-Term Price Forecasting Models Based on Artificial Neural Networks for Intraday Sessions in the Iberian Electricity Market. Energies, 9(9), 721.

[14] Muñoz, M.P., Corchero, C., Heredia, F.J. (2013). Improving electricity market price forecasting with factor models for the optimal generation bid. Int. Stat. Rev. 81(2), 289306

[15] Osório, G.J., Goncalves, J., Lujano-Rojas, J., Catalao, J. (2016). Enhanced forecasting approach for electricity market prices and wind power data series in the shortterm. Energies 9(693), 119.

[16] Murthy, G.G.P., Sedidi, V., Panda, A.K., Rath, B.N. (2014). Forecasting electricity prices in deregulated wholesale spot electricity market - a review. Int. J. Energy Econ. Policy 4(1), 32.