Hybridisation of Fundamental and Composite Econometric Modelling for Short-Term Electricity Price Forecasting

Rodrigo A. de Marcos, Antonio Bello, Javier Reneses
Institute for Research in Technology (IIT) of the ICAI School of Engineering
Comillas Pontifical University, Madrid, Spain
demarcos@comillas.edu, abello@comillas.edu, javierr@comillas.edu

Abstract**-- Traders and practitioners in diverse power exchanges are nowadays being most exposed to uncertainty than ever. This calls for appropriate electricity price forecasting models that can account for relevant aspects for electricity price forecasting. Therefore, some authors in the current literature have proposed combining fundamental models with econometric methods and have provided useful results for medium-term applications. However, little has been done for the short term. All of these facts encourage this work, which aims to the creation of a suitable hybrid short-term electricity price forecasting model, while extending the most commonly used hybridisation method by considering flexible combination approaches, several variables and individual model amalgamations. In order to validate the advantages and contributions of the proposed model, it has been applied to a real-size power exchange with complex price dynamics, such as the Iberian electricity market, and its performance was compared to that of other, more traditional, forecasting models.**

Index Terms**-- Econometric Models, Fundamental Models, Hybrid Models, Power Exchanges, Short-Term Forecasting**

I. Introduction

The ongoing evolutions that are exhibited in power exchanges have brought up a new competitive environment in which traders and practitioners must adapt their strategies and look for support for their decision-making when operating in power exchanges. This is a result of a combination of factors, which are, among many others, the increasing renewable penetration in power systems and ongoing regulatory changes. Due to the resulting uncertainty, electricity price forecasting models are frequently being resorted to for several purposes, such as risk management. These purposes are mainly framed within short-term contexts, and thus the focus of this work will be centred on forecasting horizons that are suitable for the short term.

The current literature on short-term electricity price forecasting is mostly focused on statistical or econometric approaches, e.g. time series models such as ARIMA and its variants [1], [2]; and artificial intelligence models such as neural network and support vector machine methods [3]–[5]. Many authors have combined both of these approaches, e.g. [6]–[8], in order to be able to model linear as well as non-linear trends, which are both present in electricity prices [9]. Several of these models rely on explanatory information such as demand or renewable generation forecasts.

However, econometric models are poorly capable of incorporating other relevant electricity market drivers, such as regulatory constraints or generation unit technical features. These other factors are well represented by fundamental models, which mainly aim to represent the power system, including all physical elements (generation units, transmission lines, etc.) and law-related limits (CO_2 emissions, renewable energy subsidies, taxes, etc.). Nevertheless, fundamental models are not utilised on their own due to their poor performance at yielding short-term dynamics on their electricity price estimations, which are mostly obtained by simulating the market clearing [10].

Consequently, some authors have opted to combine fundamental models with econometric approaches in order to incorporate the most relevant electricity market factors [10]–[15]. This combination has proven to be beneficial, although it has mostly been carried out in medium- and long-term contexts, and thus their adequacy for short-term forecasting horizons currently remains either poor or untested. This is therefore the main motivation of this work, which is aimed at the creation of a suitable short-term fundamental-econometric hybrid electricity price forecasting model and its application to a full-size electricity market, such as the Iberian (Spain and Portugal) power exchange.

Furthermore, one of the main reasons as to why this kind of hybrid models is avoided in short-term contexts is its extremely large size and resolution times in real power systems, mainly due to the hourly data arrangement. This issue calls for simplification methods that are usually targeted at the temporal resolution or the structure of the system, such as the generation technology aggregation that has been carried out in [13]. This simplification provides a reduction of the volume of the input data, which may also be effectively reduced by means of variable selection and shrinkage tools.

Moreover, as mentioned earlier, electricity prices are affected by a wide range of factors, and thus variable selection tools are being resorted to in order to screen out the least useful and most noisy information, such as LASSO and elastic nets [16], [17]. Another way of selecting the most relevant data is via the similar days method [18], recently applied to electricity price forecasting in [19].

It is worth noting that in this kind of hybrid models, most authors take advantage of the estimated price of the fundamental model [12]–[15], but it would be of interest to ascertain if other variables, such as unit generation, may prove more useful. Moreover, it is of interest to elaborate a composite econometric model to diversify its forecast and therefore reducing risk and volatility. The contributions of this work are summarised as follows:

- A new hourly short-term hybrid forecasting methodology is proposed and developed, which is based on a cost-production optimisation model linked to a combined neural network procedure. The usefulness of the generation outputs provided by the fundamental model was evaluated compared to the market clearing price.
- The forecasting performance of the proposed hybrid forecasting model is tested on a real-size market with complex price dynamics: the Iberian electricity market.
- Benchmark models, including well-recognised methods and individual components of the hybrid model, were tested on the same case studies in order to compare their forecasting accuracy with that of the proposed model.

The remainder of this work is organised as follows: Section II describes the proposed methodology; Section III presents the results of the experiments that were conducted with the proposed methodology as well as other models; and Section IV contains the conclusions that were drawn in this work, including the suggestions for extensions and future developments of the proposed methodology.

II. PROPOSED METHODOLOGY

The general purpose of this work is to create a new short-term hybrid electricity price forecasting model and put it to the test on a real, full-scale and complex electricity market, such as the Iberian power exchange. A general overview of the proposed model is shown on the following figure:

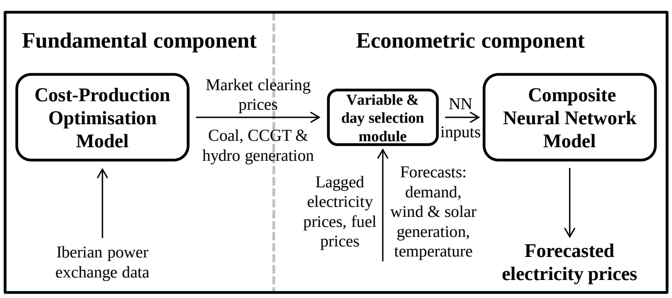

Figure 1 – Overview of the proposed fundamental-econometric model

The first part, displayed on the left side of Figure 1, is a cost-production optimisation model that has been applied to the Iberian power exchange. This model is similar to the one presented in [14], although several thermal generation units that are subject to similar fuel costs and technical constraints were grouped into larger ones (in terms of maximum power generation). By doing so, the optimisation model's runtime was reduced. Specifically, the simplified optimisation problem for a forecasting period of one week consists of 12440 equations and 71024 variables with a total runtime of 3.91 seconds and a maximum RAM usage of 76 MB on a 64-bit Windows 7 PC with 16 GB installed RAM and the following processor: Intel® Core™ i7-3770 CPU @ 3.40 GHz of 4 cores and 8 logical processors.

As shown on Figure 1, the following outputs of the optimisation model were used for the remaining parts of the proposed methodology: market clearing prices (obtained as a dual variable of the demand and generation balance constraint), coal, CCGT and hydro unit generation levels. These variables not only present relevant information for electricity price forecasting purposes, but also provide indicators of the operation and behaviour of the power exchange as well as its physical and regulatory constraints.

In order to evaluate the usefulness of these variables, they were put to the test, alongside more traditional predictors, via elastic nets [20]. These other predictors are the following: expected demand, expected solar generation, expected wind generation, coal (API2) and natural gas (NBP) month-ahead future prices. Some of these variables are available at the Spanish ISO website (https://www.esios.ree.es/en) and at the Quandl financial data platform (https://www.quandl.com).

Moreover, in order to account for seasonality and several levels of autoregression, lagged electricity market prices (1 day, 2 days, 1 week and 2 weeks) and two dummy variables (indicating if it is a business day, and holiday or Sunday) were also used as predictors. These variables (15 in total) were evaluated by means of the elastic nets method in order to remove the most redundant of them given a calibration period. By doing so, the level of complexity of the neural network is reduced and forecasting efficiency as well as algorithm parsimoniousness is increased. The calibration set data that are used for the neural network models have been arranged according to the timeline of Figure 2:

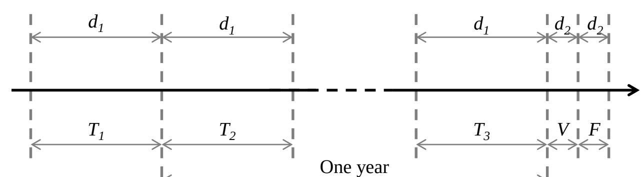

Figure 2 – Training, validation and test/forecast periods arrangement

The training set is divided into three periods of d_1 days each: T_1, T_2 and T_3. The last period is placed right before the validation period V and contains the most updated information for the neural network to train on. Periods T_1 and T_2 present relevant information pertaining to similar conditions (weather, season, etc.) in the previous year. This training set organisation is more efficient and better represents the trends in electricity prices related to the forecasting period F. Moreover, both V and F periods are set to the same length, which is of d_2 days.

Furthermore, it has been proven in [15] that the information provided by the fundamental model provides a higher overall accuracy, although it worsens it during hours of uncommonly low and high prices. This is mainly due to the effects of the market clearing prices of the fundamental model, which are somewhat centred on the daily average price level. Therefore, this calls for a combination procedure that aims to reduce forecasting error on said hours. In this case, a combination based on similar days was performed, which consists of taking the forecasts provided by a neural network model during the most similar days in order to train a second neural network model for the forecasting day. This combination procedure is depicted on Figure 3.

Given a certain forecasting day *F*, a neural network model is run for a high number of simulations and trained with the variables indicated by the elastic nets method. This model has been run for both with and without the outputs of the fundamental model. Furthermore, given *D* similar days (obtained via the similar days method), the neural networks that were trained with and without the fundamental information for day *F* are further utilised considering the data rearrangement of Figure 2 for each similar day. The forecasts of these *D* days plus day *F* were used in order to train a final neural network that provided the definitive forecast for day *F*.

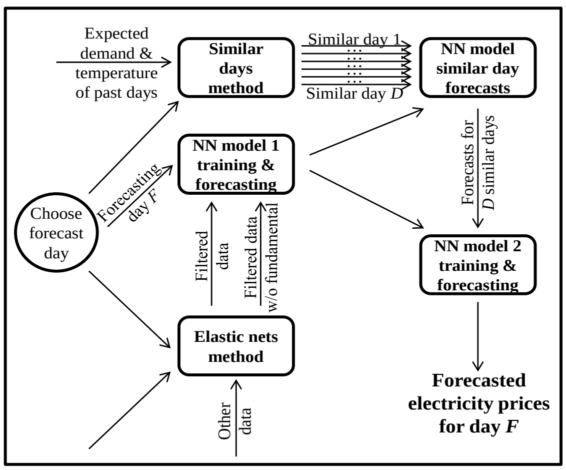

Figure 3 – Econometric component of the proposed hybrid model

The similar days method that has been applied in this work is similar to the one proposed in [18]. The expected demand (*ED*), expected demand deviation and expected temperature (*ET*) have been used in order to deem a specific day similar to the day whose price is being forecasted (day *F*). The following Euclidean norm with weighted factors was used in order to test the similarity between the day *F* that is being forecasted and another day *P* that belongs to past days:

$$\|D\| = \sqrt{D_1 + D_2 + D_3} \tag{1}$$

$$D_1 = \widehat{w}_1 (ED_t^F - ED_t^P)^2 \tag{2}$$

$$D_2 = \widehat{w}_2 [(ED_t^F - ED_{t-1}^F) - (ED_t^P - ED_{t-1}^P)]^2 \tag{3}$$

$$D_3 = \widehat{w}_3 (ET_t^F - ET_t^P)^2 \tag{4}$$

The weighted factors are ascertained by means of the following linear regression equation that is evaluated across every hour *t* that belongs to the days whose similarity is being tested:

$$ED_{t+1} = \widehat{w}_1 ED_t + \widehat{w}_2 (ED_t - ED_{t-1}) + \widehat{w}_3 ET_t \tag{5}$$

For a specific forecasting day *F*, the most similar day is that which presents the lowest value of $\|D\|$. The days, whose similarity to day *F* is evaluated, pertain to the training period T_3 (see Figure 2). Given *D* similar days, the most similar one was used as validation and the remaining *D-1* for the second neural network training.

The neural networks are trained as per the standard Levenberg-Marquardt algorithm, which is one of the most popular neural network training methods in electricity price forecasting applications, such as [3]. Several numbers of neurons in the hidden layer were considered during the training process, and the number of neurons that yielded a network with the lowest mean squared error (MSE) on a validation period forecast (*V*) was later used in order to ascertain the electricity market price for the forecasting period (*F*). The MSE is calculated as per the following formula with the conventional notation ($\hat{Y}_i$ represent the forecasted values for a certain period of *N* hours, whereas Y_i are the real values pertaining to the same period):

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{Y}_i - Y_i\right)^2 \tag{6}$$

Furthermore, a high amount of replications of the neural network forecasts are carried out in order to account for the variability and the randomness of the neural network initial weights. The final forecast is obtained by computing the mean of the obtained results in every replication. The forecasting performance is evaluated according to the mean absolute percentage error (MAPE):

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{\hat{Y}_i - Y_i}{Y_i}\right| \tag{7}$$

III. CASE STUDIES, RESULTS AND DISCUSSION

In this section, the specific details of the case studies and proposed model are presented, as well as the results and comparisons with other electricity price forecasting models.

First of all, the cost-production optimisation model has been run for all training, validation and forecast periods considered in the econometric component of the proposed hybrid model, which includes certain periods of 2016 and 2017, depending on the day *F* that is going to be forecasted and the data arrangement of Figure 2. As a result, the electricity market clearing price as well as hydro, CCGT and coal generation levels were obtained and later used for the neural network models. These variables, alongside the other predictors mentioned in the previous section, were tested with the elastic nets method and the 4 most redundant of them were removed. However, coal and natural gas prices were not considered in this test, because they are somewhat redundant if CCGT and coal generation outputs are also taken into account. These prices are only used as replacements for said generation outputs if these were not discarded by the elastic nets method for the tests without fundamental information.

Regarding the training, validation and forecasting set arrangement of Figure 2, d_1 was set to 30 days and d_2 to one day. Therefore, a total of 90 days were used as training data in order to train the neural networks for a validation period of one day and predict the electricity market price a day in advance. As for the similar days method, the 5 most similar to a specific forecasting day *F* were taken for the neural network forecast combination part of the proposed hybrid model. Moreover, the number of replications neural network training was set to 50 as the mean of the forecasts presented very little changes when reaching that amount of simulations.

Furthermore, as it is common in the literature, e.g. [19], the proposed forecasting model has been evaluated for every season of the year. Moreover, in order to further validate the proposed methodology, its forecasts have been compared with five other electricity price forecasting model benchmarks, two of which represent the two forecasts prior to the combination, i.e., the first neural network model that was trained with and without the outputs of the fundamental component (BM_1 and BM_2 respectively).

The third benchmark (BM_3) is a slight modification of a linear regression model that was recently utilised in [17] and originally introduced in [21], in which a logarithmic transformation is first applied to the electricity prices and their in-sample mean is subtracted. However, in order to account for zero-value prices in the Iberian power system, the mirror-log transform, recently applied to electricity price forecasting in [17], was applied instead of the other logarithmic transform.

The fourth benchmark (BM_4), based on ARIMA models, is well recognised and more established in the literature. The utilised model consists of a transfer function with SARIMA noise, which has been developed according to the procedures presented in the works of [22], [23]. The BIC value of the fitted models was used as model selection criterion. The obtained SARIMA noise presents the following parameters with the standard notation: $SARIMA(1,0,0)_{168}(1,0,2)_{24}(1,0,0)_1$. The expected demand was used as an exogenous variable in this model that can be therefore also referred to as a SARIMAX model.

The last benchmark (BM_5) is a simple naïve approach in which the actual electricity market prices from the previous week are directly taken as the forecast:

$$P_{day,hour} = P_{day-7,hour} \tag{8}$$

All these models have been put to the test in the Iberian power exchange for several days pertaining to all seasons of the year 2017, and the corresponding results can be seen on the following table:

TABLE I
COMPARISON OF THE PROPOSED FORECASTING MODEL WITH FIVE BENCHMARKS IN TERMS OF MAPE (%)

Model	Winter	Spring	Summer	Autumn	Average
Proposed	**8.556**	**4.266**	**3.268**	**2.938**	**4.757**
BM_1	10.417	5.099	4.361	3.400	5.819
BM_2	10.401	6.327	4.257	3.994	6.245
BM_3	17.674	11.591	6.710	4.778	10.188
BM_4	15.430	9.290	7.978	7.731	10.107
BM_5	24.716	15.349	7.679	9.109	14.213

TABLE I shows that the forecasting error is lower in all cases for the proposed model. It is worth noting that Winter, especially January 2017, presented highly uncommon circumstances in the Iberian electricity market, such as very low temperatures, very low hydro/wind generation and high natural gas prices, and thus the forecasting error on all models are higher than during the rest of the year.

One of the advantages of combining the forecasts of BM_1 and BM_2 according to Figure 3 is, as mentioned before, merging the daily average price level given by BM_1's forecast and the price peaks of BM_2's forecast.

An example can be seen on Figure 4, which shows the case for the 6th of April of 2017. Based on the similar days, the final neural network is able to reduce the bias of both BM_1 and BM_2 and thus provide a more accurate forecast. Furthermore, this may also happen even if both BM_1's and BM_2's forecasts are similar and the real value does not lie between said forecasts. This proves that the combination procedure of Figure 3 is beneficial regardless of whether BM_1 outperforms BM_2 in certain hours of the day or vice versa.

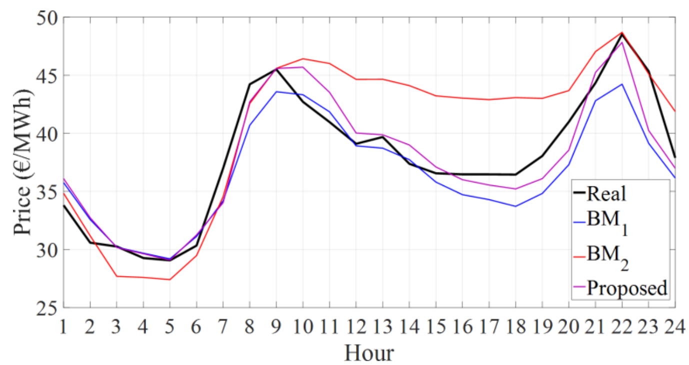

Figure 4 – Example of a forecast combination of the proposed model

Moreover, a variable selection analysis was conducted so as to test the importance of all the variables that were considered for all cases. The results of this analysis are shown on TABLE II:

TABLE II
ANALYSIS OF VARIABLE SELECTION (% OF TIMES CHOSEN)

Variable	Winter	Spring	Summer	Autumn	Average
Clearing price	100	66.7	30.8	100	74.4
Coal generation	100	75	46.2	100	80.3
CCGT generation	18.8	50	53.8	54.5	44.3
Hydro generation	18.8	8.3	30.8	63.6	30.4
Expected demand	100	100	100	100	100
Expected solar gen.	0	0	46.2	0	11.5
Expected wind gen.	87.5	100	100	100	96.9
1-day lagged prices	100	100	100	100	100
2-day lagged prices	93.8	100	100	36.4	82.5
1-week lagged prices	100	100	100	100	100
2-week lagged prices	100	100	100	100	100
Business day dummy	81.3	91.7	69.2	45.5	71.9
Sunday/holiday dummy	0	8.3	23.1	0	7.9

According to TABLE II's results, the importance of coal generation levels is slightly higher than that of market clearing prices, which may suggest that the most relevant information that is provided by the cost-optimisation model is actually coal generation levels and not market clearing prices, although in Autumn and Winter these are both equally important.

However, CCGT generation levels are the most relevant fundamental variable in the Summer season and hydro levels are mostly considered in the Autumn season, which may also provide evidence as to how the Iberian electricity market is driven by these generation technologies in every season.

IV. CONCLUSIONS AND FUTURE WORK

The novel methodology that has been proposed in this work is based on a hybrid model that is composed of a cost-production optimisation model and a composite neural network model. Both models have been linked by using some output variables of the cost-production model, such as the market clearing price, as additional inputs to the neural network models. Furthermore, the input data of the neural network models were rearranged and modified in order to increase forecasting efficiency.

The proposed hybrid model has shown adequate performance in every season of the year and the benchmark models were outperformed. Furthermore, it can be concluded that the combination of forecasts via an additional neural network and the similar days method proves beneficial.

Moreover, it is not in question that, in the short term, the importance of the market clearing prices from the fundamental model is comparable to that of the considered lagged electricity market prices. However, according to the variable selection analysis, other variables from the fundamental model may provide more useful evidence, such as coal generation levels. This should be noted when developing fundamental-econometric hybrid electricity price forecasting models in the future. Furthermore, the analysis also suggests that special attention must be paid as to how the relevance of these variables shifts between seasons.

However, even though this work's forecast combination approach proved useful, its overall computation time is roughly two times that of the individual forecast computation procedures. Therefore, other combination methods that provide faster computation times may prove useful for real short-term applications.

Additionally, the number of similar days and the number of discarded variables after testing them with elastic nets that were considered in this work were fixed to five and four respectively. These may also be modified in order to enhance this work's forecast combination approach depending on the season of the year or for every day that is being forecast, i.e. according to the training data of the neural networks, although this may increase overall computational burden, so a compromise must be achieved.

REFERENCES

[1] J. Contreras, R. Espínola, F. J. Nogales, and A. J. Conejo, "ARIMA models to predict next-day electricity prices," *IEEE Trans. Power Syst.*, vol. 18, no. 3, pp. 1014–1020, 2003.

[2] A. Cruz, A. Muñoz, J. L. Zamora, and R. Espínola, "The effect of wind generation and weekday on Spanish electricity spot price forecasting," *Electr. Power Syst. Res.*, vol. 81, no. 10, pp. 1924–1935, 2011.

[3] J. P. S. Catalão, S. J. P. S. Mariano, V. M. F. Mendes, and L. A. F. M. Ferreira, "Short-term electricity prices forecasting in a competitive market: A neural network approach," *Electr. Power Syst. Res.*, vol. 77, no. 10, pp. 1297–1304, 2007.

[4] N. Amjady, A. Daraeepour, and F. Keynia, "Day-ahead electricity price forecasting by modified relief algorithm and hybrid neural network," *IET Gener. Transm. Distrib.*, vol. 4, no. 3, p. 432, 2010.

[5] H. S. Sandhu, L. Fang, and L. Guan, "Forecasting day-ahead price spikes for the Ontario electricity market," *Electr. Power Syst. Res.*, vol. 141, pp. 450–459, 2016.

[6] J. P. S. Catalão, H. M. I. Pousinho, and V. M. F. Mendes, "Hybrid wavelet-PSO-ANFIS approach for short-term electricity prices forecasting," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 137–144, 2011.

[7] Z. Yang, L. Ce, and L. Lian, "Electricity price forecasting by a hybrid model, combining wavelet transform, ARMA and kernel-based extreme learning machine methods," *Appl. Energy*, vol. 190, pp. 291–305, 2017.

[8] S. Talari, M. Shafie-khah, G. Osório, F. Wang, A. Heidari, and J. P. S. Catalão, "Price Forecasting of Electricity Markets in the Presence of a High Penetration of Wind Power Generators," *Sustainability*, vol. 9, no. 11, p. 2065, 2017.

[9] M. Cerjan, I. Krzelj, M. Vidak, and M. Delimar, "A literature review with statistical analysis of electricity price forecasting methods," in *Eurocon, 2013 IEEE*, 2013, no. July, pp. 756–763.

[10] A. Bello, J. Reneses, A. Muñoz, and A. Delgadillo, "Probabilistic forecasting of hourly electricity prices in the medium-term using spatial interpolation techniques," *Int. J. Forecast.*, 2015.

[11] J. Nowotarski and R. Weron, "Recent advances in electricity price forecasting: A review of probabilistic forecasting," *Renew. Sustain. Energy Rev.*, 2017.

[12] A. Bello, D. Bunn, J. Reneses, and A. Muñoz, "Parametric Density Recalibration of a Fundamental Market Model to Forecast Electricity Prices," *Energies*, vol. 9, no. 11, p. 959, 2016.

[13] V. González, J. Contreras, and D. W. Bunn, "Forecasting power prices using a hybrid fundamental-econometric model," *IEEE Trans. Power Syst.*, vol. 27, no. 1, pp. 363–372, 2012.

[14] A. Bello, D. W. Bunn, J. Reneses, and A. Munoz, "Medium-Term Probabilistic Forecasting of Electricity Prices: a Hybrid Approach," *IEEE Trans. Power Syst.*, vol. 32, no. 1, pp. 334–343, 2016.

[15] R. A. de Marcos, A. Bello, and J. Reneses, "Electricity price forecasting in the short term hybridising fundamental and econometric modelling," Working Paper, 2017.

[16] A. Gianfreda and L. Grossi, "Forecasting Italian electricity zonal prices with exogenous variables," *Energy Econ.*, vol. 34, no. 6, pp. 2228–2239, 2012.

[17] B. Uniejewski, J. Nowotarski, and R. Weron, "Automated variable selection and shrinkage for day-ahead electricity price forecasting," *Energies*, vol. 9, no. 8, 2016.

[18] P. Mandal, T. Senjyu, and T. Funabashi, "Neural networks approach to forecast several hour ahead electricity prices and loads in deregulated market," *Energy Convers. Manag.*, vol. 47, no. 15–16, pp. 2128–2142, 2006.

[19] P. M. R. Bento, J. A. N. Pombo, M. R. A. Calado, and S. J. P. S. Mariano, "A bat optimized neural network and wavelet transform approach for short-term price forecasting," *Appl. Energy*, vol. 210, no. January 2018, pp. 88–97, 2018.

[20] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 67, no. 2, pp. 301–320, 2005.

[21] R. Weron and A. Misiorek, "Short-term electricity price forecasting with time series models: A review and evaluation," in *Complex Electricity Markets*, 2006, pp. 231–254.

[22] A. Pankratz, "Building Dynamic Regression Models: Model Identification," in *Forecasting with dynamic regression models (Vol. 935)*, 2012, pp. 167–201.

[23] G. Box and G. Jenkins, *Time Series Analysis - Forecasting and Control*. San Francisco: Holden Day, 1970.