

Review

# Reviewing Explanatory Methodologies of Electricity Markets: An Application to the Iberian Market

Renato Fernandes <sup>1,2,†</sup>  and Isabel Soares <sup>2,3,\*,†</sup> 

<sup>1</sup> Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência (INESC TEC), 4200-465 Porto, Portugal; renato.s.fernandes@inesctec.pt

<sup>2</sup> Faculty of Economics, University of Porto, 4200-465 Porto, Portugal

<sup>3</sup> Centro de Economia e Finanças da UP (CEFUP), 4200-465 Porto, Portugal

\* Correspondence: isoares@fep.up.pt

† These authors contributed equally to this work.

**Abstract:** In this paper, for the data set of the Iberian Electricity Market for the period 1 January 2015 to 30 June 2019, 19 different models are considered from econometrics, statistics, and artificial intelligence to explain how electricity markets work. This survey allows us to obtain a more complete, critical view of the most cited models. The machine learning models appear to be very good at selecting the best explanatory variables for the price. They provide an interesting insight into how much the price depends on each variable under a nonlinear perspective. Notwithstanding, it might be necessary to make the results understandable. Both the autoregressive models and the linear regression models can provide clear explanations for each explanatory variable, with special attention given to GARCHX and LASSO regression, which provide a cleaner linear result by removing variables that have a minimal linear impact.

**Keywords:** electricity market; machine learning; autoregressive; linear regression; GARCHX; LASSO



**Citation:** Fernandes, R.; Soares, I.

Reviewing Explanatory

Methodologies of Electricity Markets:

An Application to the Iberian Market.

*Energies* **2022**, *15*, 5020. [https://](https://doi.org/10.3390/en15145020)

[doi.org/10.3390/en15145020](https://doi.org/10.3390/en15145020)

Academic Editor: Tek Tjing Lie

Received: 31 May 2022

Accepted: 28 June 2022

Published: 9 July 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Since the EU Directives and, most particularly, over the last ten years, EU electricity markets have started shifting towards deregulation and increased competition. This had led to much uncertainty in all market agents but also to increased sophistication in the strategic behaviour of various utilities: improving efficiency either by self-improvement, often through outsourcing, or shutting down businesses with low returns while creating new and innovative businesses.

Therefore, market modelling accuracy becomes crucial, namely concerning price volatility and agents' behaviour, allowing for a timely adjustment of investment strategies and prevention/mitigation of market risks.

On the older deregulated electricity markets, the offer and supply have already stabilised. This means that even though there are still some changes in these markets, they no longer cause any drastic and disconnected ruptures but mainly cause small continuous shifts in the regular operation, depending on the gradual evolution of each market factor.

There are many contributions in the literature concerning electricity market operation and price determination, spanning through different methodologies and different datasets. It is very important to understand the current electricity market trends and to prepare for the new challenges arriving, which will potentially change the market drastically. However, it becomes difficult to forecast the impact of these new changes if, for each market, it is not possible to identify the methodology that should be used. Therefore, this paper is to consolidate these studies under a single data set to analyse the results and make transparent the advantages and disadvantages of each methodology.

In this paper, for the same data set, 19 different models are considered from econometrics, statistics, and artificial intelligence. These models were chosen for being the most cited

in the literature both from economic journals and engineering journals. The data set considered here corresponds to the Iberian Peninsula electricity market (MIBEL) from 1 January 2015 to 30 June 2019. With this setup, it is inferred which model better explains market behaviour, evaluating the model on the results readability and reliability, computation power requirements, usage difficulties and shortcomings.

## 2. Materials and Methods

This literature survey encompasses several types of models, which will be segmented by the methodology and main goals. By methodological approach considered here includes theoretical economics models, linear regression models, autoregressive models, regime-switching models, linear regression models with regularisation and machine learning models.

### Theoretical economics models

To understand how electricity spot markets work, several theoretical models were created. On the supply side, very differentiated product options are encountered, though they can be grouped into two subsets that encompass their biggest difference. There are renewable energies that produce electricity with a low marginal cost, and there are thermal energies that produce electricity at a high and variable cost, depending on the required fossil fuel. Demand in this market has been mostly inelastic so far and does not affect the market considerably. These theoretical models interpret how supply and demand meet and formulate the connections between each variable that results in market equilibrium [1]. Afterwards, these formulations are tested through econometric models. This leads to global insights into how each variable affects the market as well as qualitative implications [2,3].

The theoretical formulations mentioned above did not consider strategic behaviour from the supplier side when optimising their gains. Formulating the problem as Cournot [4–7] for price-taking firms or Bertrand [5–7] for price-setting firms, may show how different behaviours can arise. These papers formulated the market activity as a maximisation problem for multiple intervenors, resulting in either a Nash or a Pareto equilibrium. The market activity consisted of setup costs, marginal costs of productions and prices. A dynamic game was also formulated where cooperation could arise [5], as well as some static competitive games [6,7].

Within this branch of models, dispatch models can also be encountered. These models add all factual market interaction, such as merit order of the generation, supply capacity, renewable energies volatility, network capacity as well as all other variables mentioned in the previous models, but disregard strategic behaviour. Dispatch models allowed for simulations of the market and were used to determine competitive market prices [8]. With simulations of this type, it is possible to detect price deviations due to market power or other market imperfections [8]; or study the impact of a determined variable by removing it altogether from the system [9]; or even determine the level of uncertainty on parameters and decision variables [10]. Another application for the simulation of dispatch models is to study the effects of shocks in different parts of the system [11].

### Linear regression models

Linear regression models can discover a linear structure for the market interactions. Linear regressions have been used for a long time and have become the base models for many different studies. On a broader aspect, some studies used linear regressions to identify the structure of the market and analysed the intervenors behaviour. Other studies were more focused on each variable, identifying which variables affected the price and the magnitude of such effect. The market structure was identified for the US electricity and gas market [12], for German and Austrian electricity markets [13] and the European electricity markets [14]. The price effects of regulatory changes were studied for Italy [15] and Europe [16], and the effect on prices by changes in supply was studied for the European Nordic countries [17]. The impact of a single type of intervenor was studied, either through virtual power plants [18] or prosumers [19]. A descriptive study analysing the effect

of all variables against the market price was performed for Texas [20] and Europe [21]. The inertia of consumers on the choice of electricity provider was studied for Texas [22], and the price elasticity was studied for the Netherlands [23]. In all these articles, the linear regression models were obtained through the ordinary least squares and the maximum likelihood methods.

Linear regression models are defined first by their structure. Defining the data set  $\{y_i, x_{i1}, x_{i2}, \dots, x_{ip}\}_{i=1}^n$  with  $n$  observations, where  $y$  corresponds to the target variable and  $x_{i1}, \dots, x_{ip}$  corresponds to the explanatory variables. Furthermore, defining the set of coefficients  $\{\beta_0, \beta_1, \dots, \beta_p\}$  and a random error vector  $\epsilon_i$ , then it can be defined how a linear regression establishes the connection between the target variable and the explanatory variables by:

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \epsilon_i, i = 1, \dots, n \quad (1)$$

In this subsection, the linear models are solved by obtaining the set of coefficients that minimize the quadratic error. This corresponds to:

$$\min_{\beta_0, \dots, \beta_p} \sum_{i=1}^n (\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} - y_i)^2 \quad (2)$$

### Autoregressive models

Autoregressive models are linear regression models in which lags of the target variable are added to the explanatory variable pool. This type of regression is used to find time dependencies between the variables [24]. There are also direct autoregressions, in which the variable pool includes the price lags [25–31]; multivariable autoregressions, where the variable pool includes all variables, their lags and lags of the price [28,32–38]; and autoregressions with exogenous variables that are supposed to not have time dependency, i.e., the lags of some exogenous variables are not added [39,40].

The lag corresponds to the number of instants between the current target observation and a past target observation. For hourly resolution data, a 24-lag period corresponds to observing the data that occurred at the current hour in the past day. Maintaining the same definitions of the linear regression models' sub-section and also defining  $q$  as the max lags considered and  $\{\phi_1, \dots, \phi_q\}$  as the set of coefficients for the lagged variables, then the connection between the target variable and explanatory variables is defined by:

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \sum_{j=1}^q \phi_j y_{i-j} + \epsilon_i, i = 1, \dots, n \quad (3)$$

This model is solved by obtaining the set of coefficients that minimize the quadratic error. This corresponds to:

$$\min_{\beta_0, \dots, \beta_p, \phi_1, \dots, \phi_q} \sum_{i=1}^n \left( \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \sum_{j=1}^q \phi_j y_{i-j} - y_i \right)^2 \quad (4)$$

### Regime-switching models

A higher variability of models can be identified if more complexity is added to the regressions. The most frequent version within complex regressions is Markov-switching regimes, which assume there are different states in the world and that the market behaves differently in each state [41–48]. The state jump event is assumed to be random following a pre-determined distribution. Within these models, the states are supposed to be determined prior to the model usage. It is also possible to relax the latter supposition when using Hidden Semi-Markov models, in which the model itself tries to discover the number of states and which states exist [49]. Aside from Markov models, jump-diffusion models are also common, which allow for state jumps in a different way. Jump-diffusion models are a linear regression with one of the variables as a poison process, which determines the

jump event and a corresponding weight, controlling the size of the jump. The weight is usually a heavy tail probability distribution [48,50].

### Linear regression models with regularisation

If the loss function in the linear regressions is changed, then possible outcomes are the least absolute shrinkage and selection operator (LASSO) [51,52], the ridge regression [53] and the quantile regression averaging [54], which behave differently than the previously presented models. LASSO adds a penalty to the number of non-zero coefficients, leading to a lower number of explanatory variables in the model by eliminating coefficients of variables with little explanatory power [51,52]. Ridge regression adds a penalty to the sum of the coefficients, leading to coefficients closer to zero when they have little explanatory power [53]. For both these models, the connection between the target variable and the explanatory variables remains the same as:

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \epsilon_i, i = 1, \dots, n \quad (5)$$

However, the minimisation problem to be solved changes. For LASSO, it is necessary to define  $\lambda$  as a threshold parameter and  $k$  as the number of coefficients with a value different to zero. Then, the minimisation problem becomes:

$$\min_{\beta_0, \dots, \beta_p} \sum_{i=1}^n (\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} - y_i)^2 + \lambda k \quad (6)$$

For the ridge regression, it is necessary to define a magnitude parameter  $\lambda$ , and then the minimisation problem is defined as:

$$\min_{\beta_0, \dots, \beta_p} \sum_{i=1}^n (\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} - y_i)^2 + \lambda \sum_{j=0}^p \beta_j^2 \quad (7)$$

### Machine learning models

Outside of regressions scopes, it is possible to find other methods for discovering the effects of explanatory variables in the electricity price formation. The most frequently used methods are simulation models [55–57], neural networks [58–60], principal component analysis [61], singular value decomposition [62], correlation methods [63,64], gradient boosting trees [65], copula models [66] and causal determination [67]. All simulation methods surveyed were agent-based models; they correspond to models in which the intervenor actions are modelled, and their aggregated interactions produce the result. In this case, the intervenors are the energy suppliers and the energy distributors [55–57]. When creating an ensemble of regression trees, one of the possible outcomes is a gradient boosting tree, GBT for short. Overall, GBT creates thresholds within all variables, defining several paths and each path leads to a specific regression profile. With each of these paths, it is possible to better segment the electricity price to similar profiles and perform regressions on themselves [65]. Neural networks are a machine learning technique that corresponds to a set of layers that are composed of several nodes. Each node corresponds to a value and a set of weights. The layers are ordered, and each node in a layer is computed as a weighted linear combination of all nodes of the previous layer. The nodes in the first layer correspond to the explanatory variables, and the final layer, in this case, has a single node corresponding to the electricity market price. The algorithm then computes the set of weights for each node that better explains the electricity market price [58–60].

Copula models are used to model the statistic dependency between all variables in a model, which can then measure the dependency strength of each explanatory variable to the electricity market price. The copula corresponds to a multivariate cumulative distribution function. The construction of a copula is an iterative process in which each pair of variables are joined into a two-variable multivariate distribution; afterwards, each pair of two-variable multivariate distributions are joined into a three-variable multivariate

distribution and so on until all variables are joined to a single n-variable multivariate distribution [66].

Causal determination is a stronger entity than the correlation or dependency. In addition to connecting two correlated variables, it also informs which variable is the cause and which variable is the effect [67].

#### *The Iberian Electricity Market: How It Works*

The data relates to the period between 1 January 2015 and 30 June 2019 in the Iberian Electricity market, known as MIBEL. The data consist of electricity prices (price), hourly day-ahead load forecasts (demand), hourly day-ahead volatile renewable energies generation forecasts, hourly day-ahead hydric (hydro) and nuclear generation commitment, hourly day-ahead bombing (bombing) commitment, hourly day-ahead coal (coal) and combined-cycle (comb\_cycle) generation commitment, hourly day-ahead net electricity exportation to Morocco (net\_exp\_ma) and Andorra (net\_exp\_ad), weekday gas and CO<sub>2</sub> emissions market closing price, coal price on 1-month future market and weekly water reserves as generation potential energy from Portugal (reservoir\_pt) and Spain (reservoir\_sp).

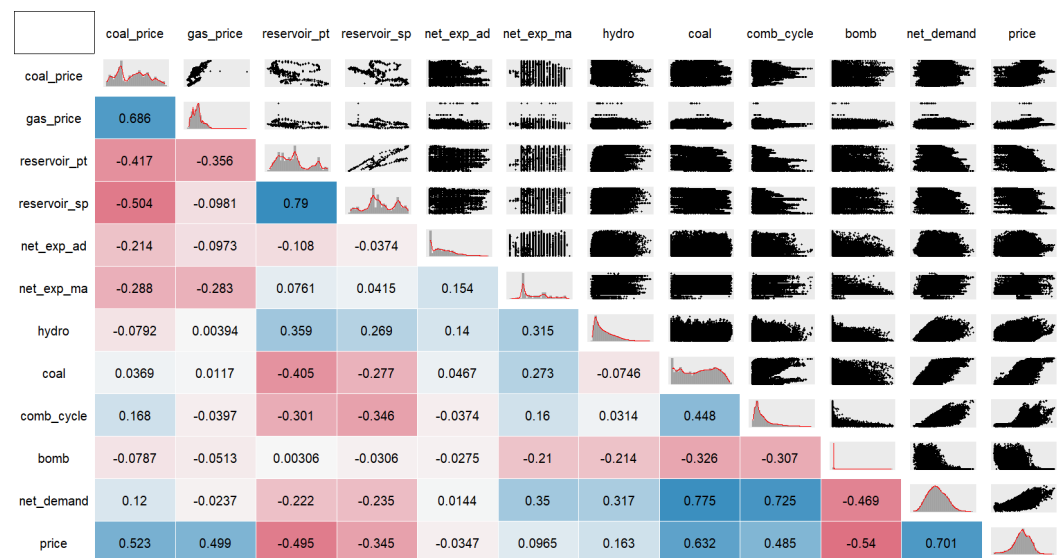
MIBEL is the Portuguese and Spanish pool market. This market works as an auction, in which bids are accepted from producers and consumers. MIBEL matches the bids maximising the welfare, which is the sum of the gains from purchase bids, sale bids and congestion charge for the 24-h period. The gain is defined as the difference between the price of the matched bid and the marginal price received. Most hourly data were obtained from OMIE, the Spanish System Operator, which included load forecasts, bombing commitments, all generation commitments, all volatile renewable energy generation forecasts and all electricity trades with countries outside the Iberian Peninsula [68]. Electricity prices and reservoir data were obtained from ENTSO-E [69]. Gas and coal prices were obtained from the Bloomberg database [70,71]. The CO<sub>2</sub> price was obtained from Sendeco2 [72].

For this study, the full scope of MIBEL was taken into consideration, and therefore, only the variables that have a direct impact on price formation will be used. This means that some variables will be dropped altogether while others will be part of a linear combination to form meaningful variables, as explained in more detail ahead. France's electricity exportation and importation are not considered because this electricity is traded as bilateral contracts, which are out of the scope of the day-ahead auction. The volatile renewable energies have a special standing by law, i.e., they enter into the market before any other generation type, and their marginal cost is close to zero. Therefore, the volatile renewable energies do not directly influence market price formulation but do influence the load that needs to be satisfied by the market. This results in creating a new variable, which will be called net\_demand and corresponds to deducting the hourly volatile renewable energies forecasts from the hourly load forecasts. Since both forecasts are volatile, which results in discrepancies between forecast and real values, joining up these two variables will result in a variable with higher volatility. One extra consideration is made in relation to the nuclear generation commitment, and due to the inherent risks of nuclear generation, there is little variation on the generation level. This corresponds to several hours of constant production and sporadic stable ramps until reaching the new level of generation. Furthermore, the generation cost of nuclear energy is close to zero, and although the generation capacity of nuclear energy is close to 20%, according to ENTSO-E data [69], the average production in the considered period is approximately 3.5% of the total generation and the maximum share of production observed was approximately 10%. Therefore, in the case of MIBEL price determination, nuclear generation has little significance, and therefore, it can also be deducted from the net\_demand previously defined. Thus, net\_demand is equal to demand minus volatile renewable energy generation and nuclear generation.

Regarding hydro, coal and combined cycle generation, MIBEL uses the merit order to decide which generation type enters the market first. Given that MIBEL is an auction-based market, by default, this is ruled by a price and quantity defined by each producer and each consumer. Using the merit order means that in MIBEL, price is a combination of the price

defined by the producer and the emission cost of CO<sub>2</sub> that will be generated to produce the electricity. In MIBEL, the average conversion rate to CO<sub>2</sub> emissions across all units that use coal is 0.9 and 0.4 for gas [65]. This results in two new variables, coal\_price and gas\_price. The variable coal\_price is equal to coal price on the 1-month future market plus 0.9 times the CO<sub>2</sub> emission market closing price. The variable gas\_price is equal to the gas price on the weekday market closing price plus 0.4 times the CO<sub>2</sub> emission market closing price.

Furthermore, MIBEL demand must be met by the supply, which means that the sum of all supply must equal the sum of all demand and bombing, excluding an always-present error margin that can be caused by multiple factors, which is solved in the intra-day market or last-minute reserve mechanisms. Figure 1 shows this through the correlation of net\_demand with coal (0.775) and cc (0.725). On the other hand, coal and combined cycle generation are not very correlated (0.448) between themselves. In terms of modelling, this generates an identification problem and, as such, models. The best option for this study is to focus on the supply side, so either net\_demand or the set of coal and combined cycle generation must be removed from the problem, as their interactions are the ones that will define the price. However, this leads to a loss of information, which can be countered on some models by sampling the observations in relation to the net\_demand quantiles.



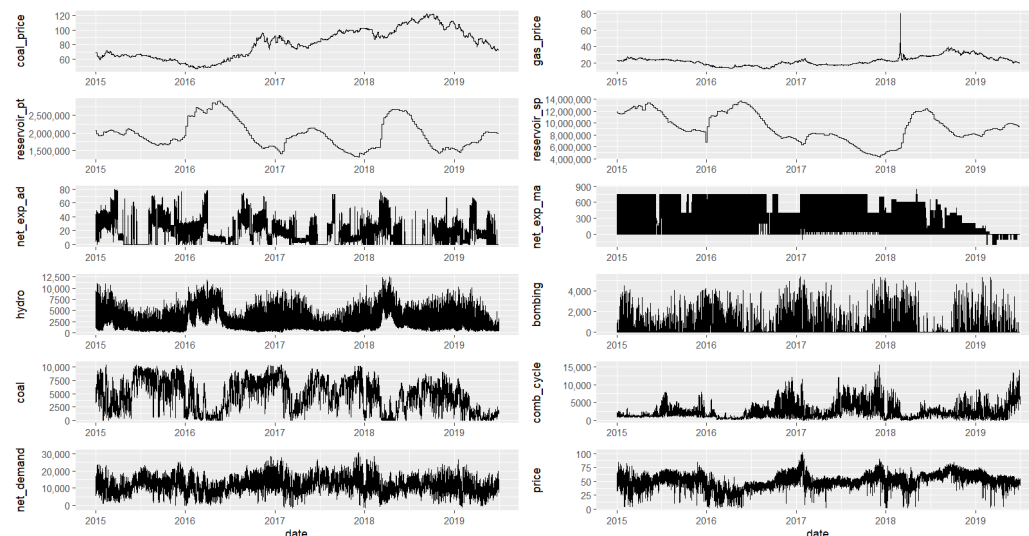
**Figure 1.** Correlation values in the lower corner, scatter plots in the upper corner and histograms with density in the diagonal for all possible time series to be used in this study.

Furthermore, the correlation between reservoir\_pt and reservoir\_sp is very high (0.79), which means adding both variables can be problematic. In relation to seasonality behaviour, both variables are very similar, as shown in Figure 2, but quantity-wise, reservoir\_pt is always smaller than reservoir\_sp. This is explained by two factors. Firstly, the hydrographic basin of Portugal is much smaller than the one in Spain. Secondly, all of the main Portuguese rivers, except for Mondego, have their origin in Spain; therefore, the main control over the river flow is in Spain. Given that hydro generation has a low marginal cost, the electricity price will be affected by when hydro generation is or is not an option. Given the lower quantity level of reservoir\_pt, this variable is the one chosen to model.

Further correlations encountered are not significant in terms of causing problems for the model performance. However, the positive correlation between coal\_price and gas\_price (0.686), coal\_price and price (0.523), gas\_price and price (0.499) is still notable. In the negative spectrum of correlations, there are coal\_price and reservoir\_pt (−0.417), reservoir\_pt and price (−0.495), bombing and net\_demand (−0.469), bombing and price (−0.54). Observing the variables histogram and distribution density function, the variables gas\_price, net\_exp\_ad, net\_exp\_ma, hydro, cc and bomb are positively skewed, with bomb

as the most skewed one. This means that these variables are mostly observed with smaller values. The net\_demand and price variable resemble a normal distribution.

Considering Figure 2, reservoir\_pt and reservoir\_sp show a clear seasonality, which was expected with the high precipitation during winter, causing the reservoirs to fill up and the hot and dry summer to empty the reservoirs. Coal generation also presents a seasonal behaviour. Furthermore, reservoir\_pt, reservoir\_sp and hydro show higher values at the beginning of the years 2016 and 2018, which was expected as those two years were wet years. This fact is also evident in the coal variable, as in those two periods, coal values are lower than the remaining years.



**Figure 2.** Time series for all possible variables to be used in this study between 1 January 2015 and 30 June 2019.

In terms of oddities, there are two strange events in this dataset. The most noticeable one is in gas\_price within the early year of 2018, with an extreme spike. This value can either be a result of some market power strategy or a simple error while filling in the data. A less noticeable one is in reservoir\_sp near the beginning of 2016, with a temporary cut in the reservoir value. In this case, a market power play makes little sense, which leads most likely to an error while filling in the database.

### 3. Results

In this section, most of the previously mentioned models will be applied to the MIBEL data. Firstly, it is important to note which models will not be applied and why. Afterwards, the analysis of the models that can be worked with will be shown.

The theoretical economy models will not be evaluated, as their results are highly and mostly influenced by the model definition itself. These models are constructed by making some assumptions on how the market works, and with these assumptions, the model is created and then the data are fed into them. After that, if the outcome is consistent with the observations, then one of the results is that the initial assumptions seem to be correct. Then, some derivations are taken from the initial assumptions to provide some more insights. Making this explanation short and bitter, most outcomes and insights come directly from the initial assumptions if the model proves itself to be reliable. Therefore, the explanation for the market using these models is not a product of the model but a product of the initial assumptions of the author. Therefore, it is not possible to analyse the explanatory power of the model but only the explanatory power of the author.

In addition to these algorithms, hidden semi-Markov models and jump-diffusion models were not studied either because these two algorithms were too sensitive to the initial parameters. Hidden semi-Markov models, depending on their random initial states,

would cause errors in the middle of the execution due to finding a near-singular matrix before performing an inverse operation. Jump-diffusion models have too many initial parameters, and the results are highly susceptible to the model's initial parameters, so there was no assurance that the results were due to the algorithm's explanatory power or the author's influence.

Agent-based simulation models can be framed in similar terms to the theoretical economic models. Agent-based simulation models require a base set of assumptions made by the author on how each agent will interact. The outcome is highly influenced by the initial assumptions made by the author.

### 3.1. Input and Output Analysis

The following methodologies will be evaluated for their usage and respective results:

1. Linear regression model,
2. Linear regression model with scaled variables,
3. Ridge regression model,
4. LASSO regression model,
5. Autoregressive model (AR),
6. Autoregressive moving averages model with exogenous variables (ARMAX),
7. Vector autoregression model (VAR),
8. Structural vector autoregression model (SVAR),
9. Generalised autoregressive conditional heteroskedasticity model (GARCH),
10. Generalised autoregressive conditional heteroskedasticity model with exogenous variables (GARCHX),
11. Gradient boosting trees model (GBT),
12. Neural network,
13. Copula model,
14. Causal model.

#### 3.1.1. Linear Regression Models

The first group of models to be presented is the first four models in the above list, corresponding to the more straightforward linear regression models presented in this paper. These models are grouped together because the type of information provided is very similar and can be easily summarised in Table 1. The values in the table correspond to the  $\beta$  coefficients of the regression in which the general form is given by  $y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \epsilon_i, i = 1, \dots, n$ . In this study, the intercept coefficient ( $\beta_0$ ) is omitted as it does not give a direct inference on the impact of each variable on the goal of explaining price. The last column is a special column for the LASSO regression model. This column explains the order in which each variable enters the model. This value should be read because the sooner the variable enters the model, the greater the impact it has on explaining price variation between observations.

**Table 1.** Coefficient values for the linear regression algorithms and LASSO variable order.

Variables	Linear Regression	Linear Regression with Standardised Variables	Ridge Regression	LASSO Regression	LASSO Order
coal_price	0.1504	2.9828	0.1413	0.1099	2
gas_price	0.6594	3.7457	0.5801	0.3331	4
reservoir_pt	−0.000006	−2.682	−0.000006	0	6
net_exp_ad	−0.0306	−0.5151	−0.0205	0	8
net_exp_ma	−0.0006	−0.1546	−0.0009	0	9
hydro	0.0015	3.1080	0.0011	0	7
bombing	−0.0051	−4.1406	−0.0045	−0.0028	3
coal	0.0019	5.8935	0.0015	0.0014	1
comb_cycle	0.0007	1.7016	0.0008	0.0001	5



From Table 1, it is noticeable that all algorithms give a very high weight to *gas\_price* and *coal\_price*. Coal has the highest coefficient when considering standardised variables. Furthermore, even though coal's coefficient is low in the LASSO regression, this variable was the first to be added to the model, which means that coal can explain much of the variance in price. LASSO also determines that *reservoir\_pt*, *net\_exp\_ad*, *net\_exp\_ma* and *hydro* add information to explain the variance of price, for which their coefficients were zero in this model. The cutting point for the zero coefficients in the LASSO regression is defined by the author, and for this case, it was defined as, when adding a new variable, the coefficient being lower than  $10^{-4}$ .

### 3.1.2. Autoregressive Models

The second group of models corresponds to the autoregressive models and are methodologies 5 to 10 in the list at the beginning of this section. The structure of these models differs significantly from each other, so they will be represented by selected sections of their regression. These selections will be justified case by case.

The AR model with 48 lags has the following result for this case study

$$\begin{aligned} price_t = & 1.1825 \times price_{t-1} - 0.2599 \times price_{t-2} + 0.0175 \times price_{t-3} - 0.0166 \times price_{t-4} \\ & + \dots + 0.1238 \times price_{t-23} + 0.2861 \times price_{t-24} - 0.3361 \times price_{t-25} + \dots \end{aligned} \quad (8)$$

The dotted sections correspond to sections with small coefficients that add minimal information to the model. The impact of the first 4 h lags and the 23 to 25 h lags is noticeable when explaining the price in the AR model. There is a special mention of the 1 h lag with a coefficient near 1.

The ARMAX model with  $p$  and  $q$  equal to 24 has the following representation:

$$\begin{aligned} price_t = & -0.0717 \times price_{t-1} - 0.042 \times price_{t-2} + \dots + 1.0273 \times \sigma_{t-1} + 0.9061 \times \sigma_{t-2} + \dots \\ & + 0.1985 \times coal\_price_t + 0.5435 \times gas\_price_t + 0.0142 \times net\_exp\_ad_t \\ & - 0.0001 \times net\_exp\_ma_t + 0.002 \times hydro_t + 0.0023 \times coal_t + 0.0007 \times comb\_cycle_t \\ & - 0.0024 \times bombing_t. \end{aligned} \quad (9)$$

The dotted section corresponds to sections with small coefficients that add little information to the model. Other than the coefficients related to errors in the past hour, the only high coefficients are related to the *coal\_price* and *gas\_price* variables. Any other explanatory variable has a very low coefficient in this model.

The VAR algorithm required some constraints when defining the model. The three variables, which are only updated in a time resolution greater than a day, have to be added with an option for the algorithm not to use any lags on them. These variables are *coal\_price*, *gas\_price* and *reservoir\_pt*. The VAR was run up to a lag of 24 h, and it resulted in:

$$\begin{aligned} price_t = & 0.0051 \times coal\_price_t + 0.0184 \times gas\_price_t + 0.00000006 \times reservoir\_pt \\ & + 0.0002 \times net\_exp\_ad_{t-1} + 0.0017 \times net\_exp\_ma_{t-1} + 0.0003 \times hydro_{t-1} \\ & - 0.0021 \times bombing_t + 0.0011 \times coal_{t-1} + 0.0004 \times comb\_cycle_{t-1} \\ & + 1.009 \times price_{t-1} + \dots \end{aligned} \quad (10)$$

Any successive lags past 1 h had a noticeable decline in the coefficient value by the order of  $10^{-1}$ . In general, this methodology attributed coefficients to all variables in all lags up to 24 h, but this resulted in small coefficients to every variable. The only exception was the coefficient for the price at 1 h lag, which had a value close to 1.

The SVAR algorithm had the same type of constraint as the VAR algorithm; although, in this case, the *coal\_price*, *gas\_price* and *reservoir\_pt* variables could not be used in any way. The resulting equation for *price* is

$$\begin{aligned} price_t = & 0.8467 \times price_{t-1} - 0.0063 \times net\_exp\_ad_{t-1} + 0.0007 \times net\_exp\_ma_{t-1} \\ & - 0.0017 \times hydro_{t-1} + 0.0017 \times bombing_{t-1} - 0.0002 \times coal_{t-1} \\ & - 0.0004 \times comb\_cycle_{t-1}. \end{aligned} \quad (11)$$

The SVAR also produced six more equations, one for each of the remaining variables that define the system of equations, in which each equation is an equation explaining each variable using 1 h lags of all other variables.

The GARCH model had a run issue in which the used library could not achieve convergence, probably due to the initial random coefficients defined by the algorithm itself. Resulting in a “False Convergence” error. The initial and ending result defined all coefficients as 0.05. To run this model properly, it would require finding an algorithm that had a different initial value or that allowed to define the initial value. The author of this paper has not found any solution so far.

The GARCHX algorithm managed to run because of the presence of the exogenous variables permitting, in this case, a convergence. The resulting model for  $p$  and  $q$  equal to 24 is

$$\begin{aligned} price_t = & 0.6824 \times price_{t-1} + 0.0369 \times price_{t-2} + 0.0115 \times price_{t-11} \\ & + 0.0137 \times price_{t-24} + 0.0271 \times \sigma_{t-3} + 0.0431 \times coal_t \\ & + 0.0202 \times comb\_cycle_t. \end{aligned} \quad (12)$$

In this case, there is no section because this algorithm considered that all other coefficients should be equal to zero. This algorithm gives very important information when setting the other coefficient to zero, meaning that the corresponding variable whose coefficients are not zero are actually very meaningful. These are the prices with lags of 1, 2, 11 and 24, which are aligned with what the AR algorithm informed. It also indicates that the errors at 3 h lags are very important, plus the current values of *coal* and *comb\_cycle*.

### 3.1.3. Machine Learning Models

The remaining models in this study are the ones in the machine learning umbrella and the last four models in the list in the beginning of this section. Starting with the neural network, the case of two hidden layers was tested. Defining  $x_i$  as the explanatory variables,  $y$  as the target variable and  $h_j^l$  as the nodes of each layer, where  $l$  corresponds to the layer and  $j$  to each corresponding node in that layer, and also defining  $w_{i,j}^l$  as the weight connecting the nodes from layer  $l - 1$  to layer  $l$ , then:

$$\begin{aligned} h_j^1 &= f_1(\sum_i x_i \times w_{i,j}^1), \\ h_j^2 &= f_2(\sum_i h_i^1 \times w_{i,j}^2), \\ y &= f_3(\sum_i h_i^2 \times w_{i,j}^3). \end{aligned} \quad (13)$$

It is important to notice that the activation functions  $f_1, f_2, f_3$  are not necessarily the same, and normally they are not linear. This means that the resulting equation for the price equation has a degree above 1, which therefore makes the resulting equation for  $y$  very hard to read and explains each variable explanatory effect on  $y$ . In addition to this main drawback of neural networks, there is another drawback, that is, the number of observations necessary to obtain convergence. Neural networks require a high volume of data, which was not achieved in this study. This means that the resulting equation changed significantly between runs. A way to solve this drawback and achieve a stable outcome is to create an assembly of neural networks, such as computing the average of

an  $N$  number of neural networks for the same data. However, on the other side, this would further increase the complexity of the model. Usually, this model is considered to be a “black box” model precisely because of this fact, as it is hard to understand the resulting model. A way to evaluate this kind of model is to fixate all variables except one and perform small modifications on the non-fixed variable. This is called a sensitivity analysis. However, given the fact there is multiplicative interaction between the variables, this analysis does not suffice to understand the model.

The causal model provides different information than any of the previously mentioned models. Instead of giving a coefficient, this model informs which variables cause which, i.e., this model says that the effect on variable A is caused by variable B, usually shortened as variable B causes variable A. The algorithm also provides extra information in the form of the order in which the connections appear, i.e., the first links are the strongest and the latter are the weakest. The ordered causality involving the price variable with the MIBEL dataset is:

- Hydro causes price;
- Coal\_price causes price;
- Gas\_price causes price;
- Bombing causes price;
- Coal causes price;
- Price causes comb\_cycle;
- Price causes reservoir\_pt.

The causality algorithm created no link between price and net\_exp\_ad or net\_exp\_ma. Although these were the results of this algorithm, it is known that bombing happens when the price is very low, not the reverse; and when gas is setting the market price, then it is known that comb\_cycle is influencing the price and not the reverse.

The GBT and Copula algorithms will be analysed together. Their direct output is very complex, but they provide a variable impact score, which is shown in Table 2. The GBT impact score evaluates the gain of each variable onto the price. Therefore, the higher this value is, the more this variable influences the model. In relation to the copula impact score, this score measures the statistical dependency of each variable with the price variable. The statistical dependency is a stronger measure than the correlation, but it can be interpreted the same as correlation. The main difference between dependency and correlation is that correlation is usually a linear relation, but dependency is not linear.

**Table 2.** Impact score of each variable under the copula and GBT algorithms.

Variables	GBT	Copula
coal_price	0.0901	−0.018
gas_price	0.1835	0.3522
reservoir_pt	0.0521	0.4779
net_exp_ad	0.0027	−0.3704
net_exp_ma	0.0107	0.2798
hydro	0.0805	0.2514
bombing	0.2573	−0.1582
coal	0.2023	0.4916
comb_cycle	0.1206	0.2753

Table 2 shows that GBT results indicate that the most important variables are bombing, coal, gas\_price and comb\_cycle. In relation the copula model, the highest positive dependencies to price are coal, reservoir\_pt and gas\_price. The highest negative one is net\_exp\_ad.

### 3.2. Computation Time Analysis

In addition to the results provided by each model and any shortcomings when applying them, it is also important to analyse how long the algorithm takes to reach its results.

Depending on the application of the models, a fast result might be more important than the quality of the results. Table 3 shows the run time of the min (fastest run), first quartile, median, third quartile and max (slowest run).

**Table 3.** Run time for each evaluated methodology in seconds.

Model	Min	1st Quartile	Median	3rd Quartile	Max
Linear Regression	0.0090	0.0092	0.0094	0.0101	0.0128
Ridge Regression	0.0049	0.0053	0.0054	0.0057	0.0186
LASSO Regression	0.0018	0.0021	0.0022	0.0023	0.0169
AR	0.0284	0.0293	0.0302	0.0325	0.1769
ARMAX	6083	6125	6162	6205	6353
VAR	9.012	9.15	9.199	9.292	10.88
SVAR	0.1162	0.1172	0.1182	0.1249	0.328
GARCH	0.1761	0.1806	0.1831	0.1856	0.2264
GARCHX	3.146	3.196	3.218	3.322	3.393
GBT	18.24	18.52	18.89	19.34	20.7
Neural Network	12,820	13,070	13150	13,240	13,690
Copula	1786	1786	1786	1787	1836
Causal	620.9	631.6	645.8	678.4	972

The group of the linear, ridge and LASSO regression are the fastest algorithms to run, with a speed on the order of  $10^{-2}$  s. The machine learning models are the slowest ones, except for GBT, and with the ARMAX algorithm with a minimum run time for this dataset above 10 min. All the other algorithms run in short by a second or by a couple of seconds. It is also worth mentioning that the first four models in Table 3 have the greatest variation between the fastest and slowest times by a factor of 10.

#### 4. Discussion

All papers identified in the Materials and Methods were used to explain electricity markets, each with their own dataset. In most papers, there is a bit of bias when presenting the preferred methodology. Every methodology has advantages and disadvantages, which is the main goal of this section.

Starting with the linear regression models, as the name indicates, these models explore the linear dependency between the explanatory variable and the price. Electricity markets cannot be fully explained by linear models. Nonetheless, linear models provide the simpler and most straightforward analysis of all models discussed in this survey. The resulting coefficients of the linear models should be interpreted as an increase of one unit of an explanatory variable that will increase or decrease the price by the same amount of the corresponding coefficient depending on the coefficient being positive or negative, respectively.

In terms of setting up linear regression models, they are very easy too. The greatest concern around these models is to avoid having the explanatory variables too heavily correlated, i.e., having a correlation factor close to 1 or  $-1$ . As for the model parametrisation, they require the set of explanatory variables, the target variable, and in the case of ridge or LASSO, they require a threshold or magnitude parameter, respectively, for the loss function. This added complexity allows for a better understanding of the model by reducing or eliminating the explanatory power of variables, which have minimal power from the beginning, making the model even easier to interpret.

The autoregressive models are highly dependent on the past history of all used variables. Depending on the autoregressive model, the explanatory variables may be only the target variable past history, or it may accept other explanatory variables too. Starting with the AR and GARCH models, these only depend on the price history to explain the price in the present. These models conceptually make little sense for the electricity market as they assume that nothing else affects the price formation other than the price in the past. That being said, these models are linear, so they are easy to understand, even though it

may be hard to fully explain some of the terms, as it is hard to explain, for example, why the last 48 h impacts the current price. It is known in electricity markets that there are specific periods on the day in which prices are higher or lower. This fact leads to accepting that the last couple hours and the hours near the 24 h mark before explaining the model. However, the rest of the terms make the model too confusing and harder to explain.

The VAR, SVAR, ARMAX and GARCHX add exogenous explanatory variables to the model. These variables can exceed the lags in the VAR and SVAR models, or just current values in the ARMAX and GARCHX. The VAR model adds all lags for all explanatory variables. For example, if 24 h lag is considered as well as the dataset explanatory variables for this model, which was six exogenous plus the price, it results in  $24 \times 7 = 168$  coefficients to understand. Furthermore, having this many coefficients dilutes the value of each coefficient.

The SVAR model has two drawbacks for electricity markets. Firstly, it fails to present information about contemporaneous events in the equation information. Secondly, it makes the assumption that the remaining variables are explained by each other as well as by the price of the hour before.

The GARCHX has similar behavior to the LASSO regression and adds multiple lags to the price and the past errors of the ARMA and GARCH algorithm. It also adds the exogenous variable's current values. However, it only presents the coefficients that have a meaningful effect on the price. This makes the GARCHX model much easier to understand and explain than the remaining presented autoregressive models.

These models also require removing highly correlated explanatory variables, but they also require giving special attention to variables that do not update hourly. If the variable does not update hourly, most of these models do not function properly if the variable is included. Furthermore, the GARCH algorithm has shown that convergence issues may arise. To set up this model, aside from the variables, it is necessary to set the number of lags and the number of past errors to be considered by the model. In terms of run time, these models stand in the middle in terms of speed, providing results in a couple of seconds, even if the result is a non-convergence error.

The machine learning models immediately bring an extra set of complexity as these models are nonlinear models. This means it is not possible to obtain direct coefficients for the explanatory variables. This increased complexity also brings about a second disadvantage, i.e., the time to run these models can ascend to several minutes in this study dataset, with the exception of the GBT model, which stood only near 20 s.

Even though these models have complex algorithms backing them up, the causal model has a very direct result. This model output states which variable directly influences the other. The order in which it provides this information is also indicative of how much confidence there is in the output. However, this outcome also shows a problem for the MIBEL dataset in that some causality directions do not agree with common knowledge on the MIBEL market, which leaves a certain distrust of the entire output. This model allows adding some rules to avoid some impossible connections, but if too many rules are created, then the outcome is defined by the author and not by the model.

Opposing the direct results of causal model, there is the neural network. This model is highly complex to the point of not being possible to obtain any type of clear impact of the explanatory variables on the price. Furthermore, for the model set up, it is necessary to indicate how many hidden layers should be in the model. A higher amount of hidden layers may or may not translate into an increase in the final model degree, and there is no straightforward way to verify if it increased or not.

The final models studied here are the GBT and the copula models. Both of their outputs are easy to analyse and indicate how strongly the explanatory variable is related to price. Furthermore, the copula model output also reveals a direction of the relation, e.g., if the explanatory variable causes an increase or decrease in the price. As for setting up the model, the GBT model has many options to be defined, e.g., to define the threshold levels, the number of random tree models to train, and learning speed to make convergence faster,

among others. In addition, it is necessary to give the target and the explanatory variables separately, and then the model is set. The copula model needs to be given all variables, including the price in the same dataset. Then, it is necessary to define the method of shaping the copula structure. In this study, it was defined as a copula centred on the price.

Overall, the model chosen to explain electricity markets is not an easy task. This survey allows obtaining a more complete view of the most cited models used to explain electricity markets. However, by analysing each model, it is possible to observe their strengths. The machine learning models appear to be very good at selecting the best explanatory variables for the price. They provide an interesting insight into how much the price depends on each variable from a nonlinear perspective. However, afterwards, it might be necessary to make the results understandable in order to reach every reader. Both the autoregressive models and the linear regression models can provide a clear explanation for each explanatory variable, with special attention to GARCHX and LASSO regression, which provide a cleaner linear result by removing variables that have a minimal linear impact. This gives a choice to the individuals who are trying to understand how the electricity market works. If there is a belief that past events up to a certain point are important, then the autoregressive models would be a good choice; or if is preferred to select the explanatory variables to introduce in the study, then the model choice would be linear regression.

**Author Contributions:** Conceptualization, I.S.; Data curation, R.F.; Investigation, R.F.; Supervision, I.S.; Writing—original draft, R.F.; Writing—review & editing, I.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Portuguese public funds through FCT—Fundação para a Ciência e a Tecnologia, I.P., in the framework of the project with reference UIDB/04105/2020.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data used in this study can be accessed from ENTSO-E [69], OMIE [68], Sendeco2 [72] and Bloomberg [70,71]. The data obtained from ENTSO-E corresponds to: Spain hourly electricity price; Spain water reservoirs and hydro storage plants; Portugal water reservoirs and hydro storage plants. The data obtained from OMIE corresponds to: Iberian peninsula load forecast; Iberian peninsula scheduled bombing commitment; Iberian peninsula scheduled generation commitment for each conventional energy type; Iberian peninsula each volatile renewable energies generation forecasts; Iberian peninsula scheduled commercial exchanges. The data obtained from Sendeco2 corresponds to: CO<sub>2</sub> prices. The data obtained from Bloomberg corresponds to: Europe Coal 6000 kcal CIF ARA Forward Month 1; Netherlands TTF Natural Gas Forward Day Ahead. The data from ENTSO-E can be freely obtained from <https://transparency.entsoe.eu/> (accessed on 1 September 2019). The data from OMIE can be freely obtained from <https://www.omie.es/en/file-access-list> (accessed on 1 September 2019) at Day-ahead Market/Total disaggregated power after Day-ahead market. The data from Sendeco2 can be freely obtained from <https://www.sendeco2.com/es/precios-co2> (accessed on 1 September 2019). The data from Bloomberg are obtained from their application, which requires a paid subscription. The full code to replicate the Results section can be found at [https://github.com/renatodsfernandes/methodology\\_study](https://github.com/renatodsfernandes/methodology_study) (accessed on 30 May 2022).

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

MIBEL	Iberian electricity market
AR	Autoregressive model
ARMAX	Autoregressive moving averages model with exogenous variables
VAR	Vector autoregression model

SVAR	Structural vector autoregression model
GARCH	Generalised autoregressive conditional heteroskedasticity model
GARCHX	Generalised autoregressive conditional heteroskedasticity model with exogenous variables
GBT	Gradient boosting trees model

## References

- Baldick, R. Electricity market equilibrium models: The effect of parametrization. *IEEE Trans. Power Syst.* **2002**, *17*, 1170–1176. [[CrossRef](#)]
- Ochoa, P. Policy changes in the Swiss electricity market: Analysis of likely market responses. *Socio-Econ. Plan. Sci.* **2007**, *41*, 336–349. [[CrossRef](#)]
- Wang, P.; Zareipour, H.; Rosehart, W.D. Characteristics of the prices of operating reserves and regulation services in competitive electricity markets. *Energy Policy* **2011**, *39*, 3210–3221. [[CrossRef](#)]
- Kamerschen, D.R.; Klein, P.G.; Porter, D.V. Market structure in the US electricity industry: A long-term perspective. *Energy Econ.* **2005**, *27*, 731–751. [[CrossRef](#)]
- Mansur, E.T. Measuring welfare in restructured electricity markets. *Rev. Econ. Stat.* **2008**, *90*, 369–386. [[CrossRef](#)]
- Newbery, D. Predicting Market Power in Wholesale Electricity Markets. *Glob. Gov.* **2009**, *RSCAS 2010*, 30. [[CrossRef](#)]
- Robinson, T.; Baniak, A. The volatility of prices in the English and Welsh electricity pool. *Appl. Econ.* **2002**, *34*, 1487–1495. [[CrossRef](#)]
- Joskow, P.; Kahn, E. A quantitative analysis of pricing behavior in California’s wholesale electricity market during summer 2000. In Proceedings of the 2001 Power Engineering Society Summer Meeting. Conference Proceedings (Cat. No.01CH37262), Vancouver, BC, Canada, 15–19 July 2001; IEEE: Piscataway, NJ, USA, 2001; Volume 3, pp. 392–394. [[CrossRef](#)]
- Dillig, M.; Jung, M.; Karl, J. The impact of renewables on electricity prices in Germany—An estimation based on historic spot prices in the years 2011–2013. *Renew. Sustain. Energy Rev.* **2016**, *57*, 7–15. [[CrossRef](#)]
- Monfared, H.J.; Ghasemi, A.; Loni, A.; Marzband, M. A hybrid price-based demand response program for the residential micro-grid. *Energy* **2019**, *185*, 274–285. [[CrossRef](#)]
- Hirth, L. What Caused the Drop in European Electricity Prices? *SSRN Electron. J.* **2016**, *39*, 143–158. [[CrossRef](#)]
- Knittel, C.R. Market Structure and the Pricing of Electricity and Natural Gas. *J. Ind. Econ.* **2003**, *51*, 167–191. [[CrossRef](#)]
- Zipp, A. The marketability of variable renewable energy in liberalized electricity markets—An empirical analysis. *Renew. Energy* **2017**, *113*, 1111–1121. [[CrossRef](#)]
- Hyland, M. Restructuring European electricity markets—A panel data analysis. *Util. Policy* **2016**, *38*, 33–42. [[CrossRef](#)]
- Bosco, B.P.; Parisio, L.P.; Pelagatti, M.M. Deregulated wholesale electricity prices in Italy: An empirical analysis. *Int. Adv. Econ. Res.* **2007**, *13*, 415–432. [[CrossRef](#)]
- Kaller, A.; Bielen, S.; Marneffe, W. The impact of regulatory quality and corruption on residential electricity prices in the context of electricity market reforms. *Energy Policy* **2018**, *123*, 514–524. [[CrossRef](#)]
- Unger, E.A.; Ulfarsson, G.F.; Gardarsson, S.M.; Matthiasson, T. A long-term analysis studying the effect of changes in the Nordic electricity supply on Danish and Finnish electricity prices. *Econ. Anal. Policy* **2017**, *56*, 37–50. [[CrossRef](#)]
- Moreno, B.; Díaz, G. The impact of virtual power plant technology composition on wholesale electricity prices: A comparative study of some European Union electricity markets. *Renew. Sustain. Energy Rev.* **2019**, *99*, 100–108. [[CrossRef](#)]
- Chesser, M.; Hanly, J.; Cassells, D.; Apergis, N. The positive feedback cycle in the electricity market: Residential solar PV adoption, electricity demand and prices. *Energy Policy* **2018**, *122*, 36–44. [[CrossRef](#)]
- Zarnikau, J.; Woo, C.K.; Zhu, S.; Tsai, C.H. Market price behavior of wholesale electricity products: Texas. *Energy Policy* **2019**, *125*, 418–428. [[CrossRef](#)]
- Pelagatti, M.M.; Bosco, B.; Parisio, L.; Baldi, F. A Robust Multivariate Long Run Analysis of European Electricity Prices. *SSRN Electron. J.* **2012**, 1–29. [[CrossRef](#)]
- Hortaçsu, A.; Madanzadeh, S.A.; Puller, S. *Power to Choose? An Analysis of Consumer Inertia in the Residential Electricity Market*; Technical Report; National Bureau of Economic Research: Cambridge, MA, USA, 2015. [[CrossRef](#)]
- Lijesen, M.G. The real-time price elasticity of electricity. *Energy Econ.* **2007**, *29*, 249–258. [[CrossRef](#)]
- Girish, G.P.; Vijayalakshmi, S. Determinants of Electricity Price in Competitive Power Market. *Int. J. Bus. Manag.* **2013**, *8*, 70–75. [[CrossRef](#)]
- Alsaedi, Y.; Tularam, G.A.; Wong, V. Application of autoregressive integrated moving average modelling for the forecasting of solar, wind, spot and options electricity prices: The Australian national electricity market. *Int. J. Energy Econ. Policy* **2019**, *9*, 263–272. [[CrossRef](#)]
- Escribano, A.; Ignacio Peña, J.; Villaplana, P. Modelling Electricity Prices: International Evidence. *Oxf. Bull. Econ. Stat.* **2011**, *73*, 622–650. [[CrossRef](#)]
- Fong Chan, K.; Gray, P. Using extreme value theory to measure value-at-risk for daily electricity spot prices. *Int. J. Forecast.* **2006**, *22*, 283–300. [[CrossRef](#)]
- García-Martos, C.; Rodríguez, J.; Sánchez, M.J. Modelling and forecasting fossil fuels, CO2 and electricity prices and their volatilities. *Appl. Energy* **2013**, *101*, 363–375. [[CrossRef](#)]

29. Koopman, S.J.; Ooms, M.; Carnero, M.A. Periodic seasonal reg-ARFIMA-GARCH models for daily electricity spot prices. *J. Am. Stat. Assoc.* **2007**, *102*, 16–27. [[CrossRef](#)]
30. Qu, H.; Duan, Q.; Niu, M. Modeling the volatility of realized volatility to improve volatility forecasts in electricity markets. *Energy Econ.* **2018**, *74*, 767–776. [[CrossRef](#)]
31. Thomas, S.; Ramiah, V.; Mitchell, H.; Heaney, R. Seasonal factors and outlier effects in rate of return on electricity spot prices in Australia's National Electricity Market. *Appl. Econ.* **2011**, *43*, 355–369. [[CrossRef](#)]
32. de Menezes, L.M.; Houllier, M.A. Reassessing the integration of European electricity markets: A fractional cointegration analysis. *Energy Econ.* **2016**, *53*, 132–150. [[CrossRef](#)]
33. García-Martos, C.; Rodríguez, J.; Sánchez, M.J. Mixed models for short-run forecasting of electricity prices: Application for the Spanish market. *IEEE Trans. Power Syst.* **2007**, *22*, 544–552. [[CrossRef](#)]
34. Gupta, I.; Bhatia, R. A Study of Long Run Relationship Between Spot and Futures Values Of Indices. *SSRN Electron. J.* **2019**, 562–569. doi: 10.2139/ssrn.3308718. [[CrossRef](#)]
35. Higgs, H. Modelling price and volatility inter-relationships in the Australian wholesale spot electricity markets. *Energy Econ.* **2009**, *31*, 748–756. doi: 10.1016/j.eneco.2009.05.003. [[CrossRef](#)]
36. Higgs, H.; Worthington, A.C. Systematic features of high-frequency volatility in Australian electricity markets: Intraday patterns, information arrival and calendar effects. *Energy J.* **2005**, *26*, 23–41. [[CrossRef](#)]
37. Malo, P.; Kanto, A. Evaluating multivariate GARCH models in the nordic electricity markets. *Commun. Stat. Simul. Comput.* **2006**, *35*, 117–148. [[CrossRef](#)]
38. Paschen, M. Dynamic analysis of the German day-ahead electricity spot market. *Energy Econ.* **2016**, *59*, 118–128. doi: 10.1016/j.eneco.2016.07.019. [[CrossRef](#)]
39. Boersen, A.; Scholtens, B. The relationship between European electricity markets and emission allowance futures prices in phase II of the EU (European Union) emission trading scheme. *Energy* **2014**, *74*, 585–594. [[CrossRef](#)]
40. Ramírez Hassan, A.; Montoya Blandón, S. Welfare gains of the poor: An endogenous Bayesian approach with spatial random effects. *Econom. Rev.* **2019**, *38*, 301–318. [[CrossRef](#)]
41. Cifter, A. Forecasting electricity price volatility with the Markov-switching GARCH model: Evidence from the Nordic electric power market. *Electr. Power Syst. Res.* **2013**, *102*, 61–67. [[CrossRef](#)]
42. Higgs, H.; Worthington, A. Stochastic price modeling of high volatility, mean-reverting, spike-prone commodities: The Australian wholesale spot electricity market. *Energy Econ.* **2008**, *30*, 3172–3185. [[CrossRef](#)]
43. Huisman, R.; Kiliç, M. A history of European electricity day-ahead prices. *Appl. Econ.* **2013**, *45*, 2683–2693. [[CrossRef](#)]
44. Kiesel, R.; Paraschiv, F. Econometric analysis of 15-minute intraday electricity prices. *Energy Econ.* **2017**, *64*, 77–90. [[CrossRef](#)]
45. de Lagarde, C.M.; Lantz, F. How renewable production depresses electricity prices: Evidence from the German market. *Energy Policy* **2018**, *117*, 263–277. [[CrossRef](#)]
46. Maryniak, P.; Trück, S.; Weron, R. Carbon pricing and electricity markets—The case of the Australian Clean Energy Bill. *Energy Econ.* **2019**, *79*, 45–58. [[CrossRef](#)]
47. Swider, D.J.; Weber, C. Extended ARMA models for estimating price developments on day-ahead electricity markets. *Electr. Power Syst. Res.* **2007**, *77*, 583–593. [[CrossRef](#)]
48. Wang, P.; Zareipour, H.; Rosehart, W.D. Descriptive models for reserve and regulation prices in competitive electricity markets. *IEEE Trans. Smart Grid* **2014**, *5*, 471–479. [[CrossRef](#)]
49. Apergis, N.; Gozgor, G.; Lau, C.K.M.; Wang, S. Decoding the Australian electricity market: New evidence from three-regime hidden semi-Markov model. *Energy Econ.* **2019**, *78*, 129–142. [[CrossRef](#)]
50. Meyer-Brandis, T.; Tankov, P. Multi-factor jump-diffusion models of electricity prices. *Int. J. Theor. Appl. Financ.* **2008**, *11*, 503–528. [[CrossRef](#)]
51. Uniejewski, B.; Nowotarski, J.; Weron, R. Automated variable selection and shrinkage for day-ahead electricity price forecasting. *Energies* **2016**, *9*, 621. [[CrossRef](#)]
52. Ziel, F. Forecasting Electricity Spot Prices Using Lasso: On Capturing the Autoregressive Intraday Structure. *IEEE Trans. Power Syst.* **2016**, *31*, 4977–4987. [[CrossRef](#)]
53. Barnes, A.K.; Balda, J.C. Sizing and economic assessment of energy Storage with real-time pricing and ancillary services. In Proceedings of the 2013 4th IEEE International Symposium on Power Electronics for Distributed Generation Systems, PEDG 2013—Conference Proceedings, Rogers, AR, USA, 8–11 July 2013. [[CrossRef](#)]
54. Maciejowska, K.; Nowotarski, J.; Weron, R. Probabilistic forecasting of electricity spot prices using Factor Quantile Regression Averaging. *Int. J. Forecast.* **2016**, *32*, 957–965. [[CrossRef](#)]
55. Bublitz, A.; Keles, D.; Fichtner, W. An analysis of the decline of electricity spot prices in Europe: Who is to blame? *Energy Policy* **2017**, *107*, 323–336. [[CrossRef](#)]
56. Lo Prete, C.; Hobbs, B.F. A cooperative game theoretic analysis of incentives for microgrids in regulated electricity markets. *Appl. Energy* **2016**, *169*, 524–541. doi: 10.1016/j.apenergy.2016.01.099. [[CrossRef](#)]
57. Sensfuß, F.; Ragwitz, M.; Genoese, M. The merit-order effect: A detailed analysis of the price effect of renewable electricity generation on spot market prices in Germany. *Energy Policy* **2008**, *36*, 3086–3094. [[CrossRef](#)]
58. Chaâbane, N. A hybrid ARFIMA and neural network model for electricity price prediction. *Int. J. Electr. Power Energy Syst.* **2014**, *55*, 187–194. [[CrossRef](#)]



59. Kaytez, F.; Taplamacioglu, M.C.; Cam, E.; Hardalac, F. Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines. *Int. J. Electr. Power Energy Syst.* **2015**, *67*, 431–438. [[CrossRef](#)]
60. Ugurlu, U.; Oksuz, I.; Tas, O. Electricity price forecasting using recurrent neural networks. *Energies* **2018**, *11*, 1255. [[CrossRef](#)]
61. Li, K.; Cursio, J.D.; Sun, Y. Principal component analysis of price fluctuation in the smart grid electricity market. *Sustainability* **2018**, *10*, 4019. [[CrossRef](#)]
62. Khoshrou, A.; Dorsman, A.B.; Pauwels, E.J. The evolution of electricity price on the German day-ahead market before and after the energy switch. *Renew. Energy* **2019**, *134*, 1–13. [[CrossRef](#)]
63. Fan, Q.; Li, D. Multifractal cross-correlation analysis in electricity spot market. *Phys. A Stat. Mech. Its Appl.* **2015**, *429*, 17–27. [[CrossRef](#)]
64. Zou, S.; Zhang, T. Multifractal Detrended Cross-Correlation Analysis of Electricity and Carbon Markets in China. *Math. Probl. Eng.* **2019**, *2019*, 9350940. [[CrossRef](#)]
65. Goncalves, C.; Ribeiro, M.; Viana, J.; Fernandes, R.; Villar, J.; Bessa, R.; Correia, G.; Sousa, J.; Mendes, V.; Nunes, A.C. Explanatory and causal analysis of the MIBEL electricity market spot price. In Proceedings of the 2019 IEEE Milan PowerTech, PowerTech 2019, Milan, Italy, 23–27 June 2019. [[CrossRef](#)]
66. Ignatieva, K.; Trück, S. Modeling spot price dependence in Australian electricity markets with applications to risk management. *Comput. Oper. Res.* **2016**, *66*, 415–433. [[CrossRef](#)]
67. Märkle-Huß, J.; Feuerriegel, S.; Neumann, D. Contract durations in the electricity market: Causal impact of 15 min trading on the EPEX SPOT market. *Energy Econ.* **2018**, *69*, 367–378. [[CrossRef](#)]
68. OMIE. Available online: [https://www.omie.es/en/file-access-list?parents%5B0%5D=/&parents%5B1%5D=Day-aheadMarket&parents%5B2%5D=2.Programmes&dir=TotaldisaggregatedpowerafterDay-aheadmarket&realdir=pdbc\\_stota](https://www.omie.es/en/file-access-list?parents%5B0%5D=/&parents%5B1%5D=Day-aheadMarket&parents%5B2%5D=2.Programmes&dir=TotaldisaggregatedpowerafterDay-aheadmarket&realdir=pdbc_stota) (accessed on 1 September 2019).
69. ENTSO-E. Transparency Platform. Available online: <https://transparency.entsoe.eu> (accessed on 1 September 2019).
70. Bloomberg, L.P. Europe Coal 6000 kcal CIF ARA Forward Month 1 (API2). Available online: <https://data.bloomberg.com> (accessed on 1 September 2019).
71. Bloomberg, L.P. Netherlands TTF Natural Gas Forward Day Ahead. Available online: <https://data.bloomberg.com> (accessed on 1 September 2019).
72. Sendeco2. CO<sub>2</sub> Prices. Available online: <https://www.sendeco2.com/es/precios-co2> (accessed on 1 September 2019).