Received July 15, 2021, accepted July 22, 2021, date of publication July 26, 2021, date of current version August 3, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3100076

Short-Term Forecasting for the Electricity Spot Prices With Extreme Values Treatment

ISMAIL SHAH[1], SHER AKBAR[1], TANZILA SABA[2], (Senior Member, IEEE), SAJID ALI[1], AND AMJAD REHMAN[2], (Senior Member, IEEE)

[1]Department of Statistics, Quaid-i-Azam University, Islamabad 45320, Pakistan
[2]Artificial Intelligence & Data Analytics Lab (AIDA), CCIS, Prince Sultan University, Riyadh 11586, Saudi Arabia

Corresponding author: Sajid Ali (sajidali.qau@hotmail.com)

This work was supported in part by the Artificial Intelligence & Data Analytics Laboratory, CCIS, Prince Sultan University.

ABSTRACT Nowadays, modeling and forecasting electricity spot prices are challenging due to their specific features, including multiple seasonalities, calendar effects, and extreme values (also known as jumps, spikes, or outliers). This study aims to provide a comprehensive analysis of electricity price forecasting by comparing several outlier filtering techniques followed by various modeling frameworks. To this end, extreme values are first treated with five different filtering techniques and are then replaced by four different outlier replacement approaches. Next, the spikes-free series is divided into deterministic and stochastic components. The deterministic component includes long-term trend, yearly and weekly seasonalities, and bank holidays and is estimated through parametric and nonparametric approaches. On the other hand, the stochastic component accounts for the short-run dynamics of the price time series and is modeled using different univariate and multivariate models. The one-day-ahead out-of-sample forecast results for the Italian Power Exchange (IPEX), obtained for a whole year, suggest that the outliers pre-filtering give a high accuracy gain. In addition, multivariate modeling for the stochastic component outperforms univariate models.

INDEX TERMS Electricity prices, forecasting, extreme values treatment, IPEX, parametric and nonparametric estimation.

I. INTRODUCTION

The liberalization of the electricity sector has transformed the structure of this sector by allowing the consumers and investors to make market entry openly. Indeed, the primary purpose of the liberalization was to motivate the public sectors to introduce a competitive environment among producers, sellers, and consumers in the electricity markets. The restructuring provides many benefits to the end-users, such as electricity at a lower cost, reliability of the power transmission system, etc. However, market participants' risk has increased due to electricity demand and prices' unique behavior. As electricity is a commodity that cannot be stored efficiently, thus requires immediate delivery to the end-user. Moreover, consumer demand varies continuously with strong seasonal and business cycle dependence, leading to multiple periodicities in electricity market data. Furthermore, the blackout effect of power transmission plants or the grid system adds randomness and complexity to the transmission system. Therefore, electricity spot prices reveal high volatility, sudden and, in general, unexpected extreme price changes known as spikes or jumps (sometimes called outliers). These extreme values explained a considerable amount of variation in the data, and hence, deleting them is not an option in electricity data modeling. An example of these extreme values is depicted in Figure 1 using hourly prices from the Italian electricity market. From this figure, the price spikes are evident for each load period, especially at the peak load periods.

In the energy literature, there are two fundamental ways to deal with spikes or jumps. The first method deals with outlier modeling that incorporates jumps or spikes in electricity prices via different modeling frameworks. For this purpose, various approaches have been suggested and used in the literature. These approaches deal with the spikes in a natural way that generally requires data mining and nonlinear modeling techniques, which often increase complexity in the estimation process and computational cost. The second approach is based on outlier filtering. Extreme values are identified as exceeding a certain fixed threshold value, and

The associate editor coordinating the review of this manuscript and approving it for publication was Akin Tascikaraoglu.

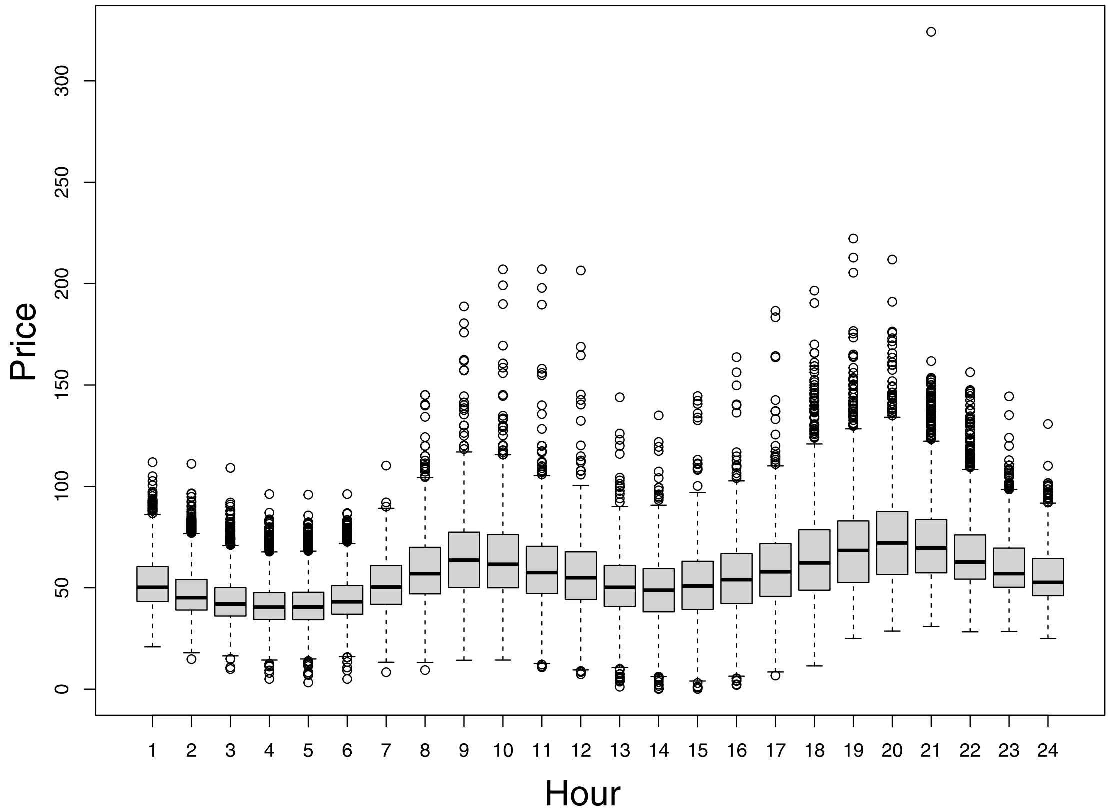

FIGURE 1. Hourly box-plots for electricity prices for the Italian power exchange (IPEX) for the period ranges from 01/01/2012 to 31/12/2016.

are replaced by a normal one to get an outlier free time series that can be modeled then. This approach uses pre-filtering techniques that are quite useful for mean price forecasting and are mostly adopted for in-sample modeling frameworks.

For electricity prices forecasting purposes, in the past, researchers used different approaches varying in complexity, methodology, and performance. For example, artificial intelligence (AI) based techniques like non-parametric regression, fuzzy logic, support vector machine and artificial neural networks (ANNs) have been used to forecast electricity prices and demand [1]–[5]. For instance, [6] studied the cascaded structure of the market clearing price (MCP) using a cascaded neural network under the Bayesian framework. The work of [2] compared the forecasting performance of different approaches, including transfer function, ARIMA, wavelet, and artificial neural networks used for electricity prices forecast. [7] and [8] also compared ANN techniques with other statistical models and concluded that ANN performs relatively better than the rest. [9] used a hybrid model containing ANNs and Hilbert-Huang transform (HHT) and conclude that it gave higher forecast accuracy than other similar techniques for electricity prices forecast. [10] studied the forecasting accuracy of electricity prices with wavelet hybridization, which involves autoregressive fractional integrated moving average (ARFIMA) along with ANNs. Results indicated a better interval forecast than other techniques used for the analysis. Recently, [11] proposed a seasonal component autoregressive artificial neural network model (SCANN) that showed high forecasting accuracy than the rest models used in the study. The performance of 10 different models, including statistical methods and computational intelligence techniques on two large European power markets, the NP and the IPEX, was investigated by [12]. The study also investigated the role of rolling window width on electricity prices forecasting performance. The authors concluded that simple models with small window sizes produce better results for peak hours, while baseload hours favor more complex specifications and longer samples.

On the other hand, statistical methods like regression and time series models are often used for electricity price and demand forecasting [13], [14]. For example, [15] studied the performance of univariate time series models for forecasting electricity demand. As the series contained daily, weekly, and annual periodicities, a multiplicative seasonal ARIMA model with double seasonal Holt-Winter, and double exponential smoothing have been used. Using regional data from IPEX, [16] study the effects of the temperature on the electricity prices formation. Using time series models, including VAR and ARMA, the results suggested that temperature has a significant explanatory power alongside traditional load variables and other structural variables. [17] studied the periodic behavior of the electricity prices using seasonal Reg-ARFIMA GARCH that explains the conditional mean and variance of electricity prices. Although the model allows for dynamic point forecasting and stochastic simulation, the results suggested that the inclusion of exogenous variables does not significantly improve the results. [18] compared the accuracy of time series in both parametric and non-parametric cases using ARIMA, SARIMA, ARCH, AR and MA models with exogenous variables ARX and ARMAX, vector AR models, threshold AR with exogenous variables TAR and TARX, regime switching model, and mean reverting jump diffusion models. The results for the point forecast indicated that models with exogenous variables performed better than the rest. Some mean reverting models have also been studied in the literature accounting for the effect of jump diffusion property of price time series [19]–[21]. For example, [22] used two techniques: the jump diffusion model and the regime switching model. The study concluded that the two modeling methods together with large number of degrees of freedom lead to better modeling performance. Negative prices can also be observed in the energy markets, and some studies showed that the regime switching model performs better when incorporating negative prices into the model [23], [24]. [25] assessed different time series modeling procedures for electricity spot prices in terms of forecasting ability. The proposed modeling procedure deals with electricity prices in two temporal segments due to the complex seasonal pattern of the electricity spot prices. The Markov switching model performs better as it captures the spiky behavior of electricity spot prices. For electricity price forecasting, [26] analyzed the accuracy of twenty-seven different models, including deep learning techniques. Their work indicated that machine learning methods yield, in general, better accuracy than statistical models. Furthermore, the study concluded that moving average terms do not improve the predictive accuracy, and hybrid models do not outperform their simpler counterparts.

Dealing with the electricity prices with high volatility, i.e., possess price spikes occurs at random in the electricity prices and thus, make it hard to be predicted and modeled. [27] studied the effect on electricity prices forecasting after excluding some of the extreme values. Generalized GARCH, dynamic regression, and exponential smoothing methods have been used to predict electricity prices with high volatility. The results suggested that accounting for extreme values can significantly improve the accuracy of the forecast. [28] studied the effect of extreme values treatment on the estimation of the seasonal and stochastic components in electricity price modeling. The study concluded that extreme observations treatment has a substantial impact on electricity price forecasting. [29] investigated the effectiveness of price spikes for risk management to the retailers, particularly focusing on the prediction of price spikes. To deal with the estimation of seasonal components and extreme values in the electricity spot prices, [30] studied different approaches. The study used different extreme values filtering techniques followed by estimating the seasonal component with different procedures. The study recommended filtering techniques before the estimation of the seasonal pattern, which gave better results.

This study aims to develop and analyze the modeling framework for forecasting electricity prices after the treatment of price spikes (extreme values). To this end, this study focuses on modeling the behavior of day-ahead electricity prices for the Italian Power Exchange (IPEX) and compared different outliers treatment techniques along with varying frameworks of forecasting to obtain an accurate day-ahead forecast. A variety of models are used to forecast electricity prices up to one-day-ahead for a whole year. Finally, the results obtained from different combinations of models are compared to get a better model in terms of forecast accuracy. Thus, the main contribution of this paper is a comprehensive investigation of different outliers filtering and replacement techniques when considering both, univariate and multivariate models for short-term electricity prices forecasting. In addition, the significance analysis of the results for the considered models is also a part of the study.

The rest of the article is organized as follows. Section II provides details about the methods used for detecting and replacing extreme values. The general modeling framework for the filtered series is given in Section III. An empirical application of the proposed methodology using the IPEX data is given in Section IV. Finally, Section V contains the conclusion and future research directions.

II. EXTREME VALUES TREATMENT TECHNIQUES

Price spike (extreme values or outliers) is defined as the price value exceeding some specific threshold value. In this study, extreme values are identified using five different identification approaches, followed by four different outlier replacement techniques. The details of these techniques are discussed below.

A. OUTLIER IDENTIFICATION TECHNIQUES

1) FIXED PRICE THRESHOLDS

Fixed Price Thresholds (TFP) technique is one of the oldest used techniques for detecting extreme values in a data set. In this method, all prices exceeding a certain level of threshold (e.g., a_0) are classified as spikes/outliers [31], [32]. Graphical techniques like sample mean excess function can be used to select a more optimal threshold value [24]. The subset containing price spikes, Y_t^o, is detected by the TFP filter as follows:

$$Y_t^0 = \{Y_t : Y_t \geq a_0\} \tag{1}$$

2) STANDARD DEVIATION FILTER ON PRICES

Based on the Standard deviation filter (SFP), the prices whose absolute deviation taken from their sample mean $\bar{Y}$ is greater than some multiple of the sample standard deviation σ. The choice of choosing a multiplier value is subjective. However, the most recommended value is 3 in the literature [20], [28], [33]. The subset containing price spikes or extreme values,

Y_t^o, is detected by the SFP as follows:

$$Y_t^0 = \{Y_t : |Y_t - \bar{Y}| \geq 3 \cdot \sigma\} \tag{2}$$

3) RECURSIVE FILTER ON PRICES

Recursive Filter on Prices (RFP) works similar to SFP, but the procedure iteratively repeats itself for the whole time series until there is no spike left. In the first step, the sample mean $\bar{Y}_1$ and the sample standard deviation σ_1 are calculated, and using equation 2, the set of spikes $Y_{t,1}^o$ are identified and replaced with the normal values. The algorithm is applied again to the obtained time series with partially removed spikes $Y_{t,1}^o$ and the spikes are replaced. This iterative procedure is repeated until no spike value is detected in the data. This procedure is extensively implemented in the literature [20], [33]. This method is also known as the variable price change threshold.

4) MOVING WINDOW FILTER ON PRICES

This technique was first proposed by [21]. Moving window filter on prices (MFP) works similar to SFP, with the only difference is that the procedure works out with a rolling window of fixed width and not for the whole time series. At the initial step, the original time series is split into $N = T/M$ parts, where M denotes the window's width and T refers to the total number of observations in the prices time series. Then, the SFP filter is applied to the first window of the given time series using equation 2. Next, the window gets shifted with M more observations ahead, and the filtering procedure is repeated till the last window. Considering that the MFP filter is a more local kind of filter than SFP filter, it is more realistic that MFP chooses fewer multiple of the standard deviations as the threshold value. In this study, Using normal distribution 95% confidence interval value (i.e., 1.96σ) is used with windows' width equal to four weeks or 672 hours. Using MFP, the set of spikes Y_t^o, with the rolling window of width M, is obtained as

$$Y_t^o = \cup_{n=1,\ldots,N}\{Y_{\tau,n} : |Y_{\tau_n} - \bar{Y}_n| \geq 1.96 \cdot \sigma_n \tag{3}$$

with $\tau_n \in ((n-1)M + 1, n \cdot M)\}$.

5) PERCENTAGE PRICE FILTER

Filtering based on the percentage price (PFP) deals with the spikes as a specific fraction of the maximum and minimum spot electricity prices. This technique was previously used by [28] and [30]. Although this technique resembles the TFP, the above mentioned fraction can be different for positive and negative spikes. Following [30], [34], we set a symmetrical threshold, i.e. 2.5%, to obtain the set of spikes/outliers Y_t^o as follows:

$$Y_t^o = \{Y_t : Y_t \leq Y_t^{\frac{2.5}{100}}\} \cup \{Y_t : Y_t \geq Y_t^{\frac{97.5}{100}}\} \tag{4}$$

where, $Y_t^{\frac{2.5}{100}}$ and $Y_t^{\frac{97.5}{100}}$ represent the percentiles of time series of orders 2.5 and 97.5, respectively.

B. OUTLIERS REPLACEMENT TECHNIQUES

Once the spikes, Y_t^o, are identified using the methods described in section II-A, they are replaced by 'normal' values. Some authors suggested to dampen prices exceeding certain threshold value by applying the logarithmic or replacing the observed price spikes with the chosen threshold value themselves [35]. [36] applied a damping scheme where all the prices above threshold Y^* were replaced by $Y_t^* = Y^* + Y^* \log_{10}(\frac{Y_t}{Y^*})$. Another way to replace the price spikes by the mean of two neighboring values [37] or replace it with one of the neighboring values [38]. However, when there are two or more consecutive spikes in the presence of multiple seasonal behavior of electricity prices, both of these replacement approaches may lead to more complications. Thus, an alternative technique was proposed by [39] where the spikes were replaced by the median of all prices with the same weekday and month as the spike has. While working with the deseasonalized data, price spikes can be replaced with the mean of deseasonalized prices. This study has used four different replacement approaches, including mean replacement, median replacement, threshold value replacement, and damping scheme approach.

III. MODELING FRAMEWORK

After possible outlier identification and replacement with the imputed values to obtain spikes free time series, the next step is to model and forecast one-day-ahead electricity prices with different modeling approaches. To this end, suppose $\boldsymbol{Y_{t,j}}$ denotes the filtered or spikes free electricity price series for $t^{th}(t = 1, 2, 3, \ldots, T)$ day and $j^{th}(j = 1, 2, 3, \ldots, 24)$ hour. The dynamics of the price series can be decomposed into deterministic $D_{t,j}$ and stochastic component $S_{t,j}$ as

$$\boldsymbol{Y_{t,j}} = D_{t,j} + S_{t,j}. \tag{5}$$

Here, the deterministic component comprised of the complex structure of the long term seasonal component (LTSC), such as, annual seasonality ($a_{t,j}$), weekly seasonality ($w_{t,j}$), bank holidays ($b_{t,j}$) and trend component ($\tau_{t,j}$). Mathematically, it can be written as

$$D_{t,j} = \tau_{t,j} + a_{t,j} + w_{t,j} + b_{t,j} \tag{6}$$

The deterministic component $D_{t,j}$ is estimated through generalized additive modeling techniques using parametric and nonparametric modeling approaches. On the other hand, the stochastic component is estimated by using four different models, including parametric autoregressive (AR), nonparametric autoregressive (NPAR), autoregressive moving average (ARMA) and vector autoregressive (VAR).

A. MODELING THE DETERMINISTIC COMPONENT

This section describes the procedure adopted for the estimation of the deterministic component. Generalized additive modeling technique is used to model the long-term trend($\tau_{t,j}$), annual ($a_{t,j}$) and weekly ($w_{t,j}$) periodicities, and bank holidays ($b_{t,j}$). The estimation of long-run (trend)

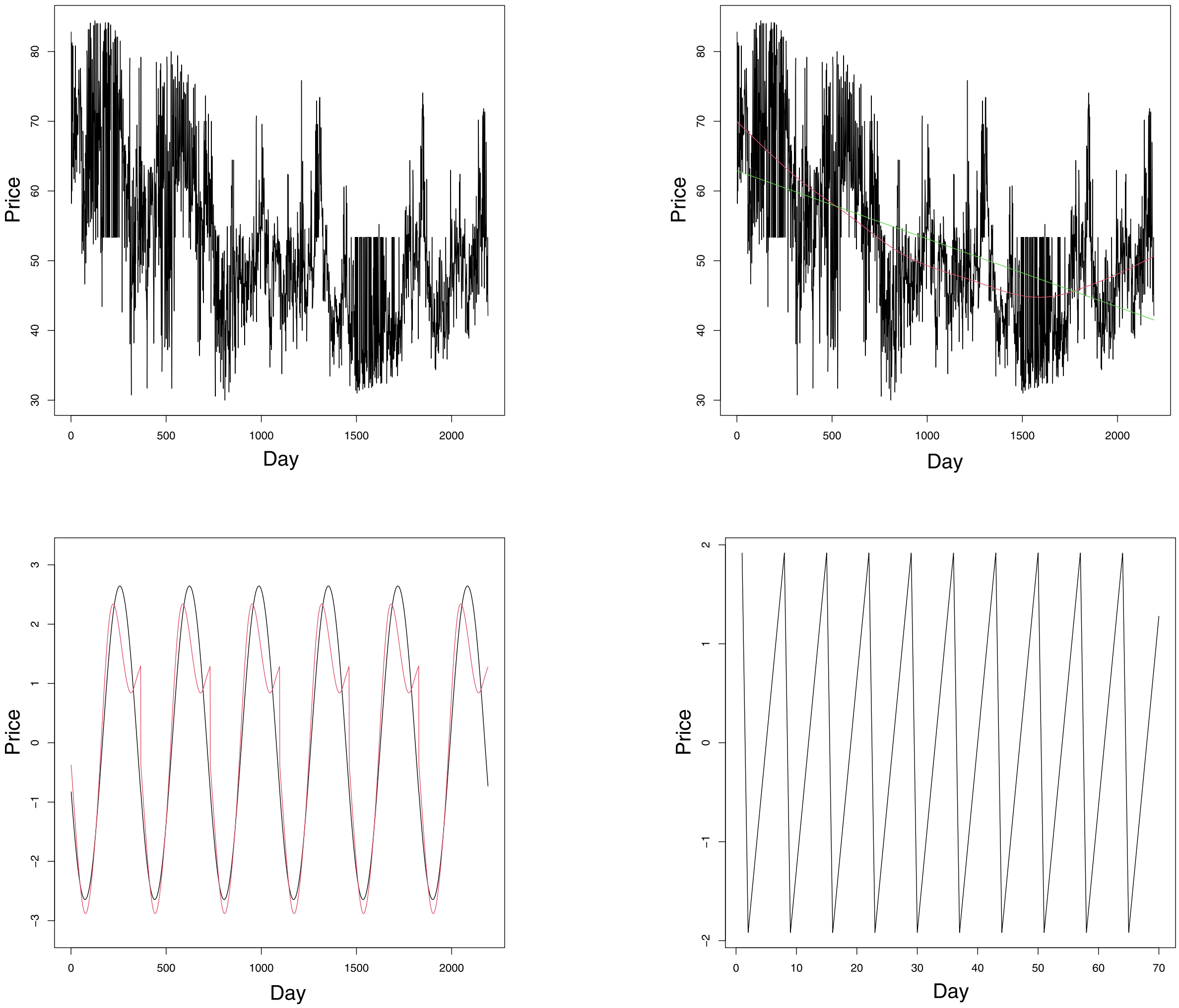

FIGURE 2. IPEX electricity prices: (Top left) filtered price time series $Y_{t,1}$, (Top right) $Y_{t,1}$ with superimposed long-trend $\tau_{t,j}$ estimated parametrically (green) and nonparametrically (red), (bottom left) yearly seasonality $a_{t,j}$ estimated parametrically (black) and nonparametrically (red), (bottom right) weekly seasonality $w_{t,j}$.

component ($\tau_{t,j}$) and annual periodicity($a_{t,j}$), which are function of time t and of the series $(1, 2, \cdots, 365, 1, 2, \cdots, 365, \cdots)$, respectively, parametric and nonparametric approaches are considered. In parametric approach, $a_{t,j}$ is modeled using sinusoidal functions, whereas $\tau_{t,j}$ by parametric regression using the ordinary least squares (OLS). In the case of nonparametric estimation, smoothing splines are used for both components to smoothly estimate them. On the other hand, dummy variables are used for the estimation of weekly ($w_{t,j}$) periodicity, and bank holidays ($b_{t,j}$) in both cases. More specifically, $w_{t,j} = \sum_{i=1}^{7} \beta_i I_{i,t}$ with $I_{i,t} = 1$ if t refers to the ith day of the week and zero otherwise. Similarly, $b_{t,j} = \sum_{i=1}^{2} \phi_i I_{i,t}$ with $I_{i,t} = 1$ if t refers to the bank holiday and zero otherwise. The coefficients β_i and ϕ are estimated with the OLS approach. An example of the estimated deterministic component is given in Figure 2. Once the deterministic components are estimated, the one-day-ahead forecast for $\hat{\tau}_{t+1,j} = \hat{\tau}_{t,j}$ and $\hat{a}_{t+1,j} = \hat{a}_{t,j}$ as both of these components represent long-term dynamics vis-à-vis our forecast horizon, whereas the forecast of $w_{t,j}$ and $b_{t,j}$ are straightforward as both of these are deterministic functions of time or the calendar conditions. Hence, one-day-ahead forecast for the deterministic component can be obtained as

$$\hat{D}_{t+1,j} = \hat{\tau}_{t,j} + \hat{a}_{t,j} + \hat{w}_{t+1,j} + \hat{b}_{t+1,j} \tag{7}$$

To obtain the stochastic component, we subtract the estimated deterministic component from $\boldsymbol{Y_{t,j}}$, i.e.,

$$\begin{aligned} \hat{S}_{t,j} &= \boldsymbol{Y_{t,j}} - \hat{D}_{t,j} \\ \hat{S}_{t,j} &= \boldsymbol{Y_{t,j}} - (\hat{\tau}_{t,j} + \hat{a}_{t,j} + \hat{w}_{t,j} + \hat{b}_{t,j}) \end{aligned} \tag{8}$$

The modeling procedure of the stochastic component given in equation 8 is discussed in the next section.

B. MODELING THE STOCHASTIC COMPONENT

This section describes the modeling and forecasting of the stochastic component. Once we obtain the stochastic component, it is modeled by a variety of parametric and nonparametric time series models. In particular, the stochastic component is modeled by using parametric autoregressive (AR), nonparametric autoregressive (NPAR), autoregressive moving average (ARMA), and vector autoregressive (VAR) models. The estimation details related to these models are as follows.

1) AUTOREGRESSIVE MODELS

The theoretical development in the field of time series analysis started early with stochastic processes. The first empirical applications of autoregressive (AR) models can be traced back to the work of George Udny Yule written in 1926. Generally, the AR process is a type of linear regression where the current value is regressed on one or more past values of the same series. Indeed, the term autoregression means the regression of the variables with itself. The AR models describe the response variable linearly depending on the most recent (lag) values and a stochastic term [40]. Autoregressive process of order p, AR(p) is written as

$$S_{t,j} = c + \alpha_1 S_{t-1,j} + \alpha_2 S_{t-2,j} + \cdots + \alpha_p S_{t-p,j} + \varepsilon_{t,j}$$

or more generally as

$$S_{t,j} = c + \sum_{i=1}^{p} \alpha_i S_{t-i,j} + \varepsilon_{t,j} \tag{9}$$

where c is the intercept term, while $\beta_i (i = 1, 2, \ldots, p)$ are the coefficients of AR(p) model. The model estimation can be done using different techniques. However, in our case, the parameters are estimated through the conditional sum of squares (CSS) method. Here it is worth mentioning that pilot analysis suggested the significant effect of lags 1,2 and 7 in most cases and thus, are used in model estimation.

2) NONPARAMETRIC AUTOREGRESIVE MODELS

The nonparametric autoregressive (NPAR) modeling approach provides a more generalized and non-linear form of estimation. NPAR accounts for the relationship between the response variable and its lagged values without considering any specific parametric form. To avoid the curse of dimensionality that refers to the exponential decay in precision as the dimension of regressors increases [41], generally, an additive form is considered. Mathematically, it can be written as

$$S_{t,j} = f(S_{t-1,j}, S_{t-2,j}, \ldots, S_{t-p,j}) + \varepsilon_{t,j}$$

or more specifically

$$S_{t,j} = f_1(S_{t-1,j}) + f_2(S_{t-2,j}) + \cdots + f_p(S_{t-p,j}) + \varepsilon_{t,j} \tag{10}$$

where, $f_j (j = 1, 2, \ldots, p)$ are smoothing functions that model the relationship between the response variable and predictors, and $\varepsilon_{t,j}$ is a disturbance term. In our case, f_j are described by cubic smoothing splines and estimated with the back-fitting algorithm. In addition, similar to the parametric case, lags 1, 2, and 7 are used in model estimation.

3) AUTOREGRESSIVE MOVING AVERAGE MODELS

In time series forecasting theory, autoregressive moving average (ARMA) is one of the most general class of models used for modeling and forecasting stationary time series data. It is a combination of AR and Moving Average (MA) models, provided that the series is stationary. As the stochastic component $S_{t,j}$ is a stationary series, the ARMA model can be used to model and forecast it. To this end, the stochastic component $S_{t,j}$ is modeled as a linear combination of the past p observations and the q lagged error terms. Mathematically, the ARMA(p,q) model can be written as

$$\begin{aligned} S_{t,j} = c + \alpha_1 s_{t-1,j} + \cdots + \alpha_p S_{t-p,j} + \epsilon_{t,j} \\ - \theta_1 \epsilon_{t-1,j} - \cdots - \theta_q \epsilon_{t-q,j} \end{aligned} \tag{11}$$

where c is the intercept term while, $\alpha_i (i = 1, 2, \ldots, p)$ and $\theta_j (j = 1, 2, \ldots, q)$ are the parameters of AR and MA components respectively, estimated by CSS method, and $\epsilon_{t,j} \sim N(0, \sigma_e^2)$. The choice of the lags of the model is an important task in ARMA models. In our case, some pilot analysis suggests to use lags 1,2, and 7 for AR and lags 1 and 7 for MA part.

4) VECTOR AUTOREGRESSIVE MODEL

The vector autoregressive model is a statistical model useful in predicting multiple time series with just one single model. The VAR is an extension of the univariate autoregressive model that extends the idea of the univariate AR model up to j time series regressions where the lag values of all j series are expressed as regressors [42]. In VAR(p) model, $S_{t,j}$ is a linear function of its own and of the other variables lags. Mathematically, VAR(p) model can be written as

$$S_t = \alpha_1 S_{t-1} + \cdots + \alpha_p S_{t-p} + \epsilon_t \tag{12}$$

Here, $\alpha_i (i = 1, 2, .., p)$ is the matrix of coefficients estimated through the CSS method, $S_t = (S_{t,1}, \ldots, S_{t,j})$ and $\epsilon_t \sim N(0, \mathbf{\Sigma}_\epsilon)$.

Once both the components, deterministic and stochastic, are estimated and forecasted, the final one-day-ahead forecasts are obtained by combining their forecasts as

$$\begin{aligned} \hat{Y}_{t+1,j} &= \hat{\tau}_{t+1,j} + \hat{a}_{t+1,j} + \hat{w}_{t+1,j} + \hat{b}_{t+1,j} + \hat{S}_{t+1,j} \\ \hat{Y}_{t+1,j} &= \hat{D}_{t+1,j} + \hat{S}_{t+1,j} \end{aligned} \tag{13}$$

The flowchart of the proposed modeling framework is given in Figure 3.

C. THE BENCHMARK MODEL

For comparison purposes, a Naïve model is also used to assess the forecasting performance of different combinations of models. This Naïve model belongs to the similar-day technique, and it proceeds as follows. To forecast, for example, Sunday, we select the day before Sunday, i.e., Saturday. We then select all the previous Saturdays in the data and compare every Saturday independently with the current Saturday price profile. The difference between any previous Saturday

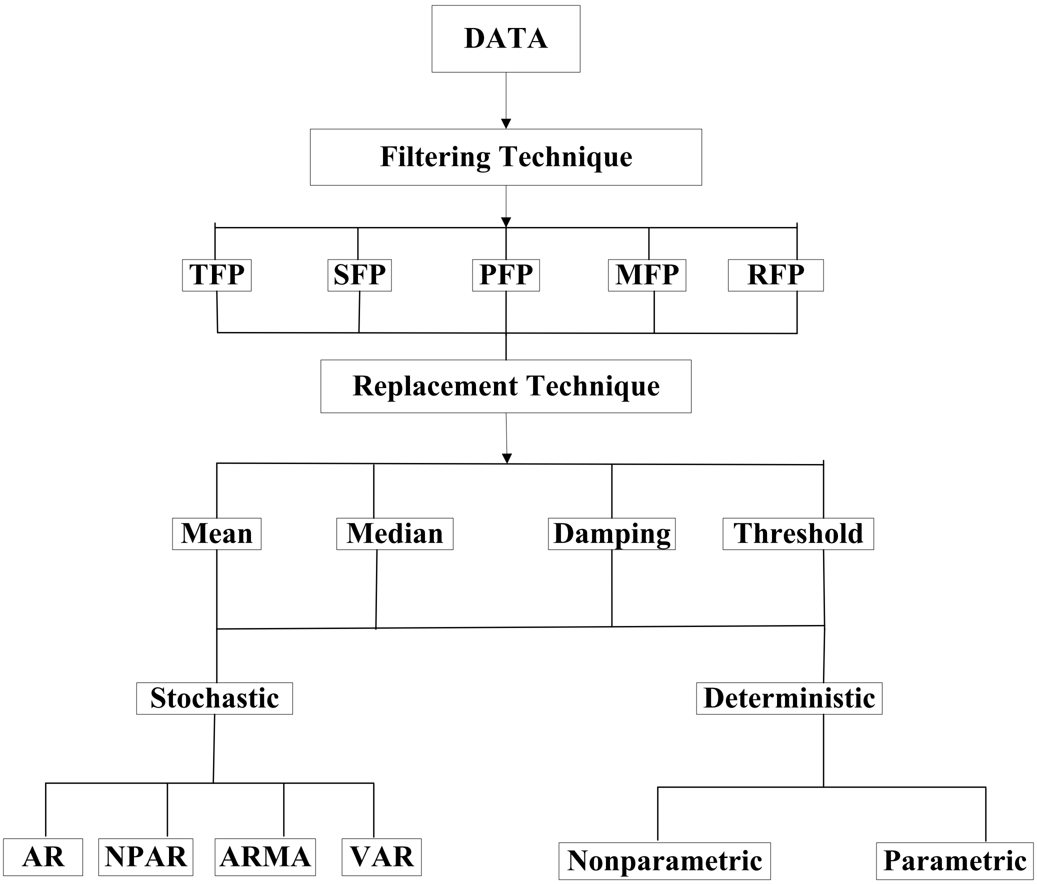

FIGURE 3. Flowchart for the proposed modeling framework.

and current Saturday is summarized using MAE (Other statistics can also be used, e.g., MAPE, however, we found MAE more useful in forecasting). In this way, we obtain a vector of MAE values for each comparison. We then find the Saturday that produces the smallest MAE value when compared to the current Saturday. Once the Saturday is identified, we use the next day to the identified Saturday, i.e., Sunday, and use it as the forecast for the Sunday we are interested in. We do the same procedure for all days of the week [43].

IV. OUT-OF-SAMPLE FORECAST FOR THE ITALIAN POWER EXCHANGE

This study considers hourly electricity spot prices data from the Italian Power Exchange (IPEX). The data set ranges from January 1^{st}, 2012 to December 31^{st}, 2017, covering 2192 days with 52608 data points. The data set is divided into two parts: from 1^{st} January 2012 to 31^{st} December 2016 (43848 data points, covering 1827 days) for the model estimation purpose, while 1^{st} January 2017 to 31^{st} December 2017 (covering 8760 data points and 365 days) for one-day-ahead post sample prediction using expending window technique. The graphical representation of the data set is given in Figure 4 where the dashed red line divided the data into model estimation and forecasting periods. To check the forecasting accuracy of the models, two standard accuracy measures, i.e., mean absolute error (MAE) and mean absolute percentage error (MAPE), are used. Mathematically, they can be written as

$$\text{MAE} = \text{mean}(|Y_{t,j} - \hat{Y}_{t,j}|)$$

and,

$$\text{MAPE} = \text{mean}(\frac{|Y_{t,j} - \hat{Y}_{t,j}|}{Y_{t,j}}) \times 100$$

where $Y_{t,j}$ is the observed (the original series) and $\hat{Y}_{t,j}$ is the forecasted one-day-ahead electricity price for $t^{th}(t = 1, 2, \ldots, 365)$ day and $j^{th}(j = 1, 2, \ldots, 24)$ hour respectively.

The forecasting results of different models are listed in Table 1 to Table 4. In each table, the results for different filtering methods with stated deterministic and stochastic models are given. The only difference among the tables is the extreme values replacement technique, which is different for each table. For example, Table 1 reports the results when we apply different filtering methods described in section II-A to original data and replaced the identified extreme values by the mean value of the series. From this table, it is evident that the series obtained through the RFP method, along with the multivariate modeling approach, gives a better forecast

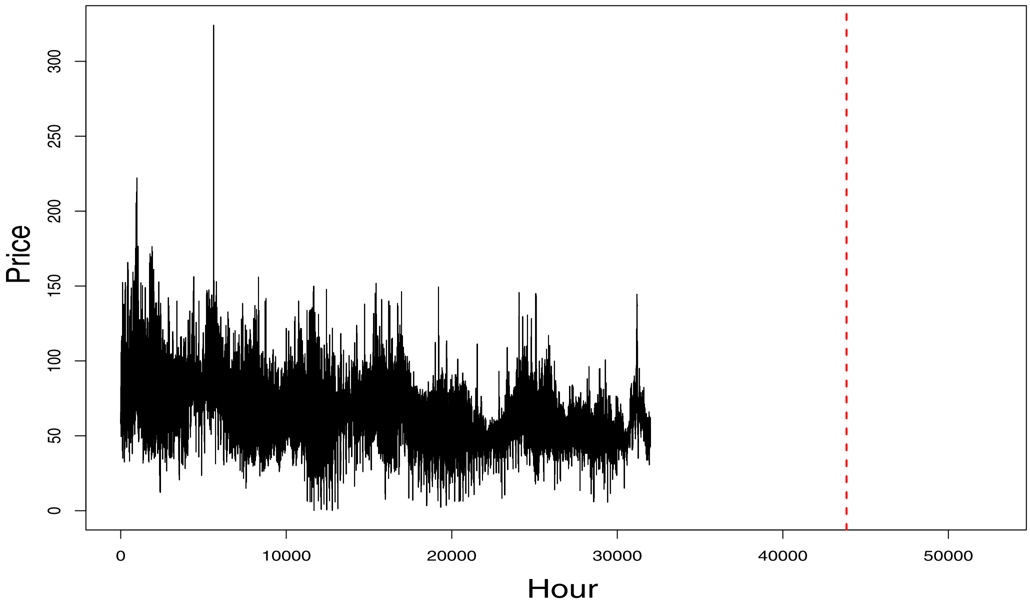

FIGURE 4. Observed hourly electricity spot prices for IPEX for the period 1^{st} January, 2012 to 31^{st} December, 2017.

TABLE 1. Mean replacement: Descriptive statistics for different combinations of models and filtering techniques. The values in bold refers to the best performing model.

Mean Replacement		NonParametric		Parametric	
Filtering Method	Models	MAE	MAPE	MAE	MAPE
TFP	AR	6.04	10.25	6.21	10.30
	NPAR	6.03	10.21	6.26	10.39
	ARMA	5.30	9.99	5.33	10.03
	VAR	5.71	9.72	5.79	9.91
SFP	AR	5.76	10.11	5.83	10.01
	NPAR	5.77	10.07	5.89	10.11
	ARMA	5.70	10.11	5.51	10.03
	VAR	5.49	9.71	5.51	9.83
PFP	AR	6.15	10.78	6.30	10.72
	NPAR	6.17	10.76	6.38	10.87
	ARMA	5.43	10.12	6.14	10.83
	VAR	5.87	10.33	5.94	10.51
MFP	AR	7.81	13.43	8.10	13.53
	NPAR	7.78	13.38	8.07	13.52
	ARMA	7.69	13.28	7.76	13.33
	VAR	7.52	12.92	7.60	12.95
RFP	AR	5.75	10.10	5.82	10.00
	NPAR	5.77	10.06	5.89	10.10
	ARMA	5.70	10.11	5.73	10.13
	VAR	**5.48**	**9.69**	**5.51**	**9.82**

TABLE 2. Median replacement: Descriptive statistics for different combinations of models and filtering techniques. The values in bold refers to the best performing model.

Median Replacement		NonParamatric		Parametric	
Filtering Method	Models	MAE	MAPE	MAE	MAPE
TFP	AR	6.05	10.28	6.25	10.34
	NPAR	6.06	10.23	6.31	10.44
	ARMA	5.96	10.24	6.01	10.28
	VAR	5.76	9.78	5.84	9.97
SFP	AR	5.77	10.12	5.85	10.03
	NPAR	5.79	10.08	5.92	10.13
	ARMA	5.72	10.13	5.74	10.15
	VAR	5.53	9.78	5.57	9.89
PFP	AR	6.16	10.75	6.32	10.71
	NPAR	6.18	10.72	6.41	10.87
	ARMA	6.10	10.72	6.15	10.80
	VAR	5.89	10.31	5.96	10.49
MFP	AR	7.81	13.34	8.12	13.47
	NPAR	7.79	13.29	8.09	13.45
	ARMA	7.70	13.20	7.77	12.74
	VAR	7.59	12.97	7.63	12.93
RFP	AR	5.77	10.12	5.85	10.03
	NPAR	5.79	10.08	5.92	10.13
	ARMA	5.71	10.12	5.74	10.15
	VAR	**5.50**	**9.75**	**5.53**	**9.88**

in both parametric and nonparametric cases. In this case, we obtain the least values for the MAE and MAPE of 5.48 and 9.69, respectively. Comparing the univariate and multivariate approaches, note that the VAR model performs better than the univariate models. Concerning the results of parametric and nonparametric estimation in the case of deterministic component, the results suggested that nonparametric estimation leads to better forecasting. Finally, the filters MFP and PFP give poor results in terms of predictive performance compared to the rest. The benchmark model, the Naïve model, is applied directly to the original price time series, and it produced a MAPE and MAE of 12.54 and 6.84, respectively, which is considerably higher than all the listed results in Table 1.

Table 2 refers to the forecasting results when the median value of the series replaces the identified extreme values. From the table, the RFP and SFP filters, along with multivariate model VAR give better forecasting accuracy compared to the rest. In contrast, the MFP and PFP filters give higher forecasting errors in terms of MAE and MAPE.

TABLE 3. Damping scheme replacement: Descriptive statistics for different combinations of models and filtering techniques. The values in bold refers to the best performing model.

Damping Scheme Replacement		NonParamatric		Parametric	
Filtering Method	Models	MAE	MAPE	MAE	MAPE
TFP	AR	5.53	9.96	5.55	9.84
	NPAR	5.50	9.90	5.53	9.82
	ARMA	5.48	9.96	5.49	9.95
	VAR	5.30	9.59	5.32	9.67
SFP	AR	5.62	9.99	5.65	9.87
	NPAR	5.63	9.98	5.69	9.94
	ARMA	5.56	9.97	5.57	9.98
	VAR	5.26	9.41	5.28	9.53
PFP	AR	5.61	10.16	5.62	9.99
	NPAR	5.53	10.01	5.54	9.89
	ARMA	5.55	10.13	5.58	10.17
	VAR	5.29	9.63	5.31	9.70
MFP	AR	5.65	10.23	5.67	10.06
	NPAR	5.59	10.10	5.59	9.99
	ARMA	5.60	10.22	5.63	10.24
	VAR	5.31	9.65	5.32	9.66
RFP	AR	5.62	9.96	5.67	9.85
	NPAR	5.64	9.94	5.71	9.93
	AIMA	5.55	9.91	5.56	9.93
	VAR	**5.25**	**9.35**	**5.27**	**9.47**

TABLE 4. Threshold value replacement: Descriptive statistics for different combinations of models and filtering techniques. The values in bold refers to the best performing model.

Threshold Value Replacement		NonParamatric		Parametric	
Filtering Method	Models	MAE	MAPE	MAE	MAPE
TFP	AR	5.53	9.96	5.55	9.85
	NPAR	5.51	9.93	5.54	9.86
	ARMA	5.47	9.94	5.48	9.94
	VAR	5.29	9.56	5.35	9.64
SFP	AR	5.53	9.97	5.55	9.83
	NPAR	5.51	9.91	5.54	9.85
	ARMA	5.47	9.93	5.48	9.94
	VAR	5.29	9.55	5.32	9.64
PFP	AR	5.53	9.93	5.54	9.80
	NPAR	5.52	9.91	5.55	9.86
	ARMA	5.44	9.86	5.46	9.87
	VAR	5.22	9.42	5.23	9.50
MFP	AR	5.60	9.90	5.64	9.82
	NPAR	5.59	9.87	5.65	9.87
	ARMA	5.50	9.82	5.51	9.82
	VAR	**5.16**	**9.05**	**5.26**	**9.40**
RFP	AR	5.53	9.95	5.56	9.83
	NPAR	5.51	9.91	5.54	9.84
	ARMA	5.47	9.93	5.48	9.94
	VAR	5.28	9.54	5.31	9.63

The nonparametric estimation of the deterministic term again leads to better forecasting results compared to the parametric case. Once again, the superior performance of the multivariate model is evident compared to the univariate models.

Table 3 elaborates the predictive performance of different models when the extreme values are replaced with damping scheme values. Results from this table show a better predictive performance of using the RFP filter along with the multivariate VAR model, both in parametric and nonparametric cases. The obtained MAE and MAPE values of 5.25 and 9.35, respectively, are considerably lower than the results listed in Table 1 and 2. Again, nonparametric estimation of

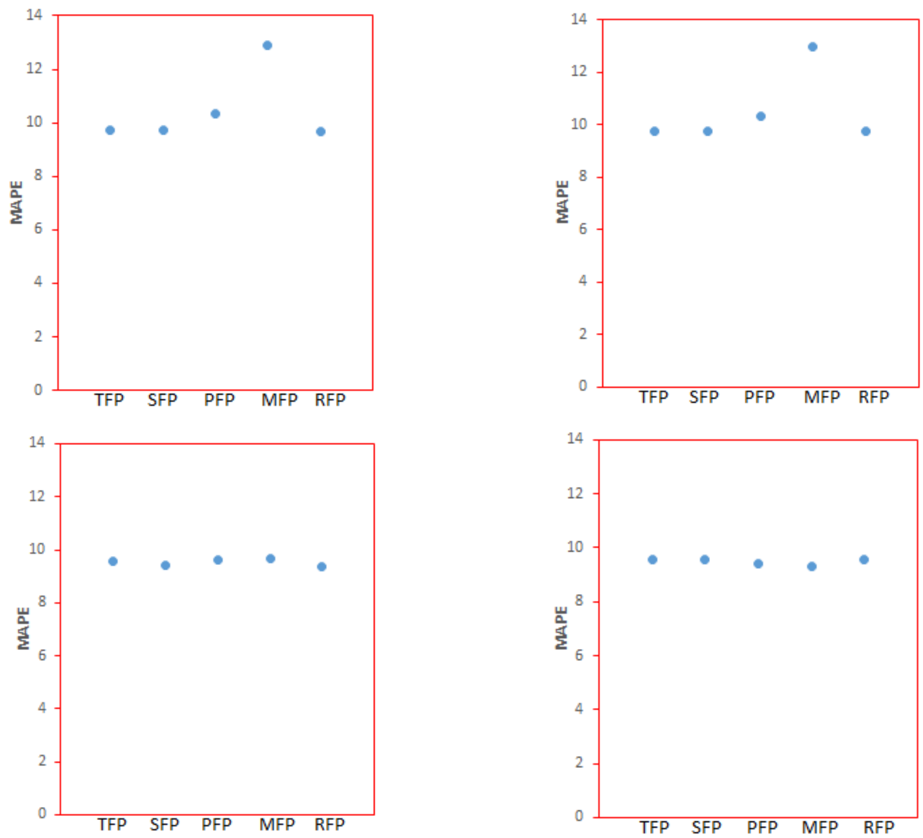

FIGURE 5. Day-ahead out-of-sample MAPE for VAR model with different extreme values filtering and replacement approaches (top left) mean replacement (top right) median replacement (bottom left) damping scheme replacement (bottom right) threshold value replacement.

deterministic component and estimating the stochastic component by a multivariate model leads to better forecasting results.

Finally, Table 4 listed the results when the extreme values are replaced by a threshold values approach. In this case, we can see that this replacement approach produces the smallest MAE and MAPE values of 5.16 and 9.05, respectively, compared with the previous tables. In addition, the MFP filter along with the multivariate VAR modeling procedure outperforms other filters in both parametric and nonparametric cases. Note that VAR combined with different extreme values replacement techniques perform relatively better than the univariate models. Due to the superior performance, the results of VAR models, combined with different filtering and replacement techniques, are depicted in Figure 5. The graphical representation of MAPE for various filtering and replacement methods suggests that damping scheme and threshold value approaches produce better results than the mean and median replacement approaches.

Looking at the Tables 1 to 4, it is evident that the lowest MAPE and MAE values are produced by MFP filter with threshold replacement technique and using nonparametric estimation for the deterministic component. Thus, to assess the significance of the differences among these results, the Diebold and Mariano (DM) [44] test is performed only for these results. The null hypothesis of the DM test corresponds to equal forecast accuracy, while the alternative hypothesis states that the model in the row is more accurate than in the column. The DM results listed in Table 5 indicate that the VAR model is statistically significant than all other models. In the case of univariate models, the difference in their results is not statistically significant. Moreover, the proposed models are statistically significant than the naïve model.

TABLE 5. P-values for the DM test. Null hypothesis: equal prediction accuracy, alternative hypothesis: model in the row is more accurate than model in the column (squared loss function used).

	AR	NPAR	ARMA	VAR	Naïve
AR	-	0.54	0.59	> 0.99	< 0.01
NPAR	0.46	-	0.48	> 0.99	< 0.01
ARMA	0.41	0.52	-	> 0.99	< 0.01
VAR	< 0.01	< 0.01	< 0.01	-	< 0.01
Naïve	> 0.99	> 0.99	> 0.99	> 0.99	-

It is worth mentioning that the predictive performance of our models is highly competitive with the results reported in the literature. For example, [43] reported a MAPE value of 9.74 obtained with the NPAR model which is higher than our obtained value of 9.05. The work of [47] reported an MAE of 8.58, whereas we obtained a value of 5.16 with the MFP-VAR.

V. CONCLUSION

This study revisited the problem of one-day-ahead electricity price forecasting in the liberalized electricity market. As the price series is highly volatile, this study particularly focused on filtering the series for extreme values. The filtered series is then modeled by using different deterministic and stochastic models. More precisely, extreme values filtering approaches include fixed price threshold (TFP), standard deviation filter (SFP), percentage price filter (PFP), moving window filter (MFP), and recursive filter on prices (RFP). The identified extreme values are replaced by four different replacement methods: mean replacement, median replacement, damping scheme, and threshold value replacement. Once the spikes-free series is obtained, the price time series is divided into two major components: deterministic and stochastic. The deterministic component includes the estimation of a long-term trend, yearly and weekly seasonalities, and bank holidays through parametric and nonparametric approaches. On the other hand, the stochastic component accounts for the short-run dynamics of the price time series. It is modeled by parametric autoregressive (PAR), nonparametric autoregressive (NPAR), autoregressive moving average (ARMA), and vector autoregressive (VAR) models. For comparison purposes, a naïve model was also computed by using the original price time series. To assess the predictive performance of the models, mean absolute error (MAE) and mean absolute percentage error (MAPE) are calculated. Data from the Italian Power Exchange (IPEX) are used and one-day-ahead out-of-sample forecasts are obtained for a complete year.

The study's empirical findings suggested that the RFP and MFP filtering techniques combined with the damping scheme and threshold value replacement generally produce lower error values. In addition, the multivariate model VAR outperformed the univariate models in all cases. Furthermore, the nonparametric estimation of the deterministic component leads to better forecasting than in the parametric case. Moreover, all the proposed models outperform the naïve model used in the study. As this study considers only one electricity market, IPEX, empirical analysis conducted on other electricity markets is recommended in the future. In addition, the effect of including exogenous variables in the models can be investigated.

ACKNOWLEDGMENT

The authors would like to acknowledge the support of Prince Sultan University for paying the Article Processing Charges (APC) of this publication.

FUNDING

No specific funding received for this research.

REFERENCES

[1] 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, pp. 88–97, Jan. 2018.

[2] A. J. Conejo, J. Contreras, R. Espínola, and M. A. Plazas, "Forecasting electricity prices for a day-ahead pool-based electric energy market," *Int. J. Forecasting*, vol. 21, no. 3, pp. 435–462, Jul. 2005.

[3] A. Y. Alanis, "Electricity prices forecasting using artificial neural networks," *IEEE Latin Amer. Trans.*, vol. 16, no. 1, pp. 105–111, Jan. 2018.

[4] X. Zhang, J. Wang, and Y. Gao, "A hybrid short-term electricity price forecasting framework: Cuckoo search-based feature selection with singular spectrum analysis and SVM," *Energy Econ.*, vol. 81, pp. 899–913, Jun. 2019.

[5] I. Shah, "Modeling and forecasting electricity market variables," Ph.D. dissertation, Dept. Stat. Sci., Univ. Padua, Padua, Italy, 2016.

[6] L. Zhang, P. B. Luh, and K. Kasiviswanathan, "Energy clearing price prediction and confidence interval estimation with cascaded neural networks," *IEEE Trans. Power Syst.*, vol. 18, no. 1, pp. 99–105, Feb. 2003.

[7] R. Pino, J. Parreno, A. Gomez, and P. Priore, "Forecasting next-day price of electricity in the Spanish energy market using artificial neural networks," *Eng. Appl. Artif. Intell.*, vol. 21, no. 1, pp. 53–62, Feb. 2008.

[8] V. Vahidinasab, S. Jadid, and A. Kazemi, "Day-ahead price forecasting in restructured power systems using artificial neural networks," *Electr. Power Syst. Res.*, vol. 78, no. 8, pp. 1332–1342, Aug. 2008.

[9] V. Kurbatsky and N. Tomin, "Forecasting prices in the liberalized electricity market using the hybrid models," in *Proc. IEEE Int. Energy Conf.*, Dec. 2010, pp. 363–368.

[10] F. Saâdaoui and H. Rabbouch, "A wavelet-based hybrid neural network for short-term electricity prices forecasting," *Artif. Intell. Rev.*, vol. 52, no. 1, pp. 649–669, Jun. 2019.

[11] G. Marcjasz, B. Uniejewski, and R. Weron, "On the importance of the long-term seasonal component in day-ahead electricity price forecasting with NARX neural networks," *Int. J. Forecasting*, vol. 35, no. 4, pp. 1520–1532, Oct. 2019.

[12] C. Fezzi and L. Mosetti, "Size matters: Estimation sample length and electricity price forecasting accuracy," *Energy J.*, vol. 41, no. 4, Oct. 2020.

[13] F. Lisi and I. Shah, "Forecasting next-day electricity demand and prices based on functional models," *Energy Syst.*, vol. 11, no. 4, pp. 947–979, Nov. 2020.

[14] I. Shah, H. Iftikhar, S. Ali, and D. Wang, "Short-term electricity demand forecasting using components estimation technique," *Energies*, vol. 12, no. 13, p. 2532, 2019.

[15] J. W. Taylor, "Short-term electricity demand forecasting using double seasonal exponential smoothing," *J. Oper. Res. Soc.*, vol. 54, no. 8, pp. 799–805, 2003.

[16] S. Bigerna, "Estimating temperature effects on the Italian electricity market," *Energy Policy*, vol. 118, pp. 257–269, Jul. 2018.

[17] S. J. Koopman, M. Ooms, and M. A. Carnero, "Periodic seasonal Reg-ARFIMA–GARCH models for daily electricity spot prices," *J. Amer. Stat. Assoc.*, vol. 102, no. 477, pp. 16–27, 2007.

[18] R. Weron and A. Misiorek, "Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models," *Int. J. Forecasting*, vol. 24, no. 4, pp. 744–763, Oct. 2008.

[19] M. Bierbrauer, S. Trück, and R. Weron, *Modeling Electricity Prices With Regime Switching Models* (Lecture Notes in Computer Science), vol. 3039. 2004, pp. 859–867.

[20] Á. Cartea and M. G. Figueroa, "Pricing in electricity markets: A mean reverting jump diffusion model with seasonality," *Appl. Math. Finance*, vol. 12, no. 4, pp. 313–335, 2005.

[21] S. Borovkova and F. J. Permana, "Modelling electricity prices by the potential jump-diffusion," in *Stochastic Finance*. Boston, MA, USA: Springer, 2006, pp. 239–263.

[22] R. Weron, M. Bierbrauer, and S. Trück, "Modeling electricity prices: Jump diffusion and regime switching," *Phys. A, Stat. Mech. Appl.*, vol. 336, nos. 1–2, pp. 39–48, May 2004.

[23] D. Keles, M. Genoese, D. Möst, and W. Fichtner, "Comparison of extended mean-reversion and time series models for electricity spot price simulation considering negative prices," *Energy Econ.*, vol. 34, no. 4, pp. 1012–1032, Jul. 2012.

[24] E. Fanone, A. Gamba, and M. Prokopczuk, "The case of negative day-ahead electricity prices," *Energy Econ.*, vol. 35, pp. 22–34, Jan. 2013.

[25] M. Bessec, J. Fouquau, and S. Meritet, "Forecasting electricity spot prices using time-series models with a double temporal segmentation," *Appl. Econ.*, vol. 48, no. 5, pp. 361–378, Jan. 2016.

[26] J. Lago, F. De Ridder, and B. De Schutter, "Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms," *Appl. Energy*, vol. 221, pp. 386–405, Jul. 2018.

[27] H. S. Guirguis and F. A. Felder, "Further advances in forecasting day-ahead electricity prices using time series models," *KIEE Int. Trans. Power Eng.*, vol. 4, no. 3, pp. 159–166, 2004.
[28] S. Trueck, R. Weron, and R. Wolff, "Outlier treatment and robust approaches for modeling electricity spot prices," in *Proc. 56th Session ISI*, Lisbon, Portugal, Aug. 2007.
[29] T. M. Christensen, A. S. Hurn, and K. A. Lindsay, "Forecasting spikes in electricity prices," *Int. J. Forecasting*, vol. 28, no. 2, pp. 400–411, Apr. 2012.
[30] J. Janczura, S. Trück, R. Weron, and R. C. Wolff, "Identifying spikes and seasonal components in electricity spot price data: A guide to robust modeling," *Energy Econ.*, vol. 38, pp. 96–110, Jul. 2013.
[31] A. Boogert and D. Dupont, "When supply meets demand: The case of hourly spot electricity prices," *IEEE Trans. Power Syst.*, vol. 23, no. 2, pp. 389–398, May 2008.
[32] C. Lapuerta and B. Moselle, *Recommendations for the Dutch Electricity Market*. London, U.K.: The Brattle Group, pp. 1–40, 2001.
[33] L. Clewlow and C. Strickland, *Energy Derivatives: Pricing and Risk Management*. Sirolo AN, Italy: Lacima Publ., 2000.
[34] D. O. Afanasyev and E. A. Fedorova, "On the impact of outlier filtering on the electricity price forecasting accuracy," *Appl. Energy*, vol. 236, pp. 196–210, Feb. 2019.
[35] M. Shahidehpour, H. Yamin, and Z. Li, *Market Operations in Electric Power Systems: Forecasting, Scheduling, and Risk Management*. Hoboken, NJ, USA: Wiley, 2003.
[36] J. Nowotarski, J. Tomczyk, and R. Weron, "Robust estimation and forecasting of the long-term seasonal component of electricity spot prices," *Energy Econ.*, vol. 39, pp. 13–27, Sep. 2013.
[37] R. Weron, *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*, vol. 403. Chichester, U.K.: Wiley, 2007.
[38] H. Geman and A. Roncoroni, "Understanding the fine structure of electricity prices," *J. Bus.*, vol. 79, no. 3, pp. 1225–1262, 2006. [Online]. Available: https://EconPapers.repec.org/RePEc:ucp:jnlbus:v:79:y:2006:i:3:p:1225-1262
[39] M. Bierbrauer, C. Menn, S. T. Rachev, and S. Trück, "Spot and derivative pricing in the EEX power market," *J. Banking Finance*, vol. 31, no. 11, pp. 3462–3485, Nov. 2007.
[40] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*. Berlin, Germany: Springer, 2006.
[41] L. Wasserman, *All of Nonparametric Statistics*. Berlin, Germany: Springer, 2006.
[42] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*. Melbourne, VIC, Australia: OTexts, 2018.
[43] I. Shah, H. Bibi, S. Ali, L. Wang, and Z. Yue, "Forecasting one-day-ahead electricity prices for Italian electricity market using parametric and nonparametric approaches," *IEEE Access*, vol. 8, pp. 123104–123113, 2020.
[44] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *J. Bus. Econ. Stat.*, vol. 13, no. 3, pp. 253–263, 1995.
[45] A. Petrella and S. Sapio, "A time series analysis of day-ahead prices on the Italian power exchange," in *Proc. 6th Int. Conf. Eur. Energy Market*, May 2009, pp. 1–6.
[46] A. Cervone, E. Santini, S. Teodori, and D. Z. Romito, "Electricity price forecast: A comparison of different models to evaluate the single national price in the Italian energy exchange market," *Int. J. Energy Econ. Policy*, vol. 4, no. 4, pp. 744–758, 2014.
[47] A. Petrella and S. Sapio, "No PUN intended: A time series analysis of the Italian day-ahead electricity prices," EUI RSCAS, Loyola de Palacio, Programme Energy Policy, Brussels, Belgium, Tech. Rep. 2010/03, 2010.

ISMAIL SHAH received the master's degree from Lund University, Sweden, and the Ph.D. degree from the University of Padua, Italy. He is currently working as an Assistant Professor with the Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan. His research interests include functional data analysis, time series analysis, regression analysis, energy economics, and applied and industrial statistics. He is also working as an Editor for the *Journal of Quantitative Methods*.

SHER AKBAR received the M.Phil. degree in statistics from Quaid-i-Azam University (QAU), Islamabad, Pakistan. His research interests include time series forecasting, energy economics, and applied statistics.

TANZILA SABA (Senior Member, IEEE) received the Ph.D. degree in document information security and management from the Faculty of Computing, Universiti Teknologi Malaysia (UTM), Malaysia, in 2012. She is currently serving as an Associate Professor and the Associate Chair of the Information Systems Department, College of Computer and Information Sciences, Prince Sultan University, Riyadh, Saudi Arabia. She has over 100 publications that have around 1800 citations with H-index 28. Her most publications are in the biomedical research, published in ISI/SCIE indexed. Her primary research interests include medical imaging, pattern recognition, data mining, MRI analysis, and soft-computing. She received the Best Student Award from the Faculty of Computing, UTM, in 2012. Due to her excellent research achievement, she is included in Marquis Who's Who (S & T) 2012. She is the Leader of the Artificial Intelligence and Data Analytics Research Laboratory, PSU, and an active Professional Member of ACM, AIS, and IAENG organizations. She is the PSU WiDS (Women in Data Science) Ambassador at Stanford University. She is currently an editor and a reviewer of reputed journals and on the panel of TPC of International conferences. On the accreditation side, she is a skilled lady with ABET and NCAAA quality assurance.

SAJID ALI received the Ph.D. degree in statistics from Bocconi University, Milan, Italy. He is currently an Assistant Professor at the Department of Statistics, Quaid-i-Azam University (QAU), Islamabad, Pakistan. His research interests include time series analysis, Bayesian inference, construction of new flexible probability distributions, and process monitoring.

AMJAD REHMAN (Senior Member, IEEE) received the Ph.D. and postdoctoral degrees (Hons.) from the Faculty of Computing, Universiti Teknologi Malaysia, with a specialization in forensic documents analysis and security, in 2010 and 2011, respectively. He is a Senior Researcher at the Artificial Intelligence & Data Analytics Laboratory, CCIS, Prince Sultan University, Riyadh, Saudi Arabia. He is currently a PI in several funded projects and also completed projects funded from MoHE, Malaysia, and Saudi Arabia. He is the author of more than 200 ISI, Scopus journal articles, and conferences. His research interests include big data mining, health informatics, pattern recognition, and forecasting. He received the Rector's Award 2010 for best student in the university.