IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 27, 2020, accepted July 1, 2020, date of publication July 6, 2020, date of current version July 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3007189

Forecasting One-Day-Ahead Electricity Prices for Italian Electricity Market Using Parametric and Nonparametric Approaches

ISMAIL SHAH[1], HAJRA BIBI[1], SAJID ALI[1], LICHEN WANG[2], AND ZHEN YUE[2]
[1]Department of Statistics, Quaid-i-Azam University, Islamabad 45320, Pakistan
[2]College of Life Science, Linyi University, Linyi 276000, China

Corresponding authors: Lichen Wang (wanglichen@lyu.edu.cn) and Zhen Yue (yuezhen@lyu.edu.cn)

The work of Lichen Wang and Zhen Yue was supported by the Program for Shandong Provincial Natural Science Foundation of China under Grant ZR2018BC012.

ABSTRACT Over the last three decades, accurate modeling and forecasting of electricity prices has become a key issue in competitive electricity markets. As electricity price series usually exhibit several complex features, such as high volatility, seasonality, calendar effect, non-stationarity, non-linearity and mean reversion, price forecasting is not a trivial task. However, participants of electricity market need price forecast to make decisions in their daily activity in the market, such as trading, risk management or future planning. In this study we consider linear and nonlinear models for one-day-ahead forecast of electricity prices using components estimation techniques. This approach requires to filter out the structural, deterministic components from the original time series and to model the residual component by means of some stochastic process. The final forecast is obtained by combining the predictions of both these components. In this work, linear and non-linear models are applied to both, deterministic and stochastic, components. In the case of stochastic component, AutoRegressive, Nonparametric AutoRegressive, Functional AutoRegressive, and Nonparametric Functional AutoRegressive have been considered. Furthermore, two naïve benchmarks are applied directly to the price time series and their results are compared with our proposed models. An application of the proposed methodology is presented for the Italian electricity market (IPEX). Our analysis suggests that, in terms of Mean Absolute Error, Mean Absolute Percentage Error, and Pearson correlation coefficient, best results are obtained when deterministic component is estimated by using parametric approach. Further, Functional AutoRegressive model performs relatively better than the rest while Nonparametric AutoRegressive is highly competitive.

INDEX TERMS Electricity prices forecasting, parametric and nonparametric models, univariate and multivariate time series, modeling and forecasting, IPEX.

I. INTRODUCTION

With the liberalization of the electricity sector, electricity has become a tradable commodity. However, due to its natural and physical characteristics, electricity as a commodity differs inherently from other commodities. For example, electricity has no shelf life because it cannot be stored in large amounts over a long period of time. It's storage possibilities are limited and expensive. Therefore, to assure stability of the power system, the amount of power fed into the grid must constantly match demand. Because of these reasons the electricity prices are highly volatile, which causes high risks to market participants [1], [2]. In addition, other factors such as weather conditions and calendar effects, may also cause the electricity price to change. For example, due to the change in climate conditions such as temperature and the number of daylight hours, electricity demand shows seasonal fluctuations which translate into seasonal behavior of electricity prices, especially in spot prices. In general, electricity prices across the world usually exhibit three different types of seasonal patterns including, daily, weekly and yearly cycles and often contain a linear or nonlinear trend. Moreover, electricity spot prices are highly volatile in nature and prices over a short period of time can vary extremely. In fact, electricity prices are far more volatile than any other financial commodity. Thus, accurate price

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

forecasts are crucial for both market participants and market operators [3]–[6].

Electricity price forecasting plays an important role in the scheduling and management of electricity markets and therefore, electricity market participants may be interested in different price forecasting horizons. Generally, electricity price forecasting can be divided into three main time horizons: short-term, medium-term and long-term price forecast. The forecasting based on these horizons is required for different purposes. For example, short-term price forecast (STPF) is usually referred to the time horizon of a few hours to a few days. STPF plays an essential role in the power scheduling and management, risk assessment and other decision making. Market participants need STPF for making better bidding strategies in order to get maximum profits in the day-ahead markets [3], [7]. Medium-term price forecasts (MTPF) range from a week to a few months and is generally required for generation expansion planning, maintenance scheduling, bilateral contacting, fuel contacting, developing investment and hedging strategies [8]. On the other hand, long-term price forecast (LTPF) generally covers time from several months to a few years and is generally used for planning and investment profitability analysis, such as making decisions for future investments in power plants, inducing sites and fuel sources [9]. As in many electricity markets daily prices are determined the day before the physical delivery by means of hourly concurrent auction, STPF has received higher attention in the literature.

Electricity price forecasting is much complex than forecasting electricity demand because of its unique characteristics, uncertainties in operations as well as the bidding strategies of the market participants [2]. In the past, extensive studies have been made on the problem of electricity prices forecasting using different modeling techniques and procedures. Statistical models such as time series models, regression models, and exponential smoothing methods are widely used to forecast electricity market variables. Statistical models generally use a mathematical combination of the past and current information of the endogenous variable (prices) with, sometimes, exogenous variables included to the model. Autoregressive (AR), Vector autoregressive (VAR), Autoregressive moving average (ARMA) model and its different extensions like seasonal ARIMA, ARIMA with exogenous variables (ARIMAX), autoregressive conditional heteroskedasticity (ARCH) and generalized autoregressive conditional heteroskedasticity (GARCH) models are some of the common time series methods used for electricity price forecasting problems in the literature [10]–[14]. For example, for the Spanish and California electricity market, Contreras, *et al.* [10] used ARIMA models to predict hourly electricity prices. For both markets, their models produced reasonable errors compared to the Artificial Neural Networks. Liu and Shi [15] used 10 ARMA–GARCH approaches to model and forecast for hourly electricity prices. Using electricity prices time series from the New England electricity market, they noted that the ARMA–GARCH(-M) models are effective for modeling the electricity prices. Regression models are the most commonly used statistical techniques. Regression models, e.g., multiple regression models are used to learn the relationship between dependent (interest) variable and other independent variables. In our specific context, they are used to model the relationship of current electricity prices with other influential factors such as demand, calendar conditions, temperature, fuel prices, etc. Regression models are usually easy to implement and interpret and are widely used for electricity prices forecasting [16]–[19]. For example, based on multiple linear regression analysis, Ferreira, *et al.* [18] investigated the impact of various explanatory variables on the electricity price modeling for the Portugal and Spain electricity markets. They concluded that multiple linear regression model is a plausible strategy to obtain causal forecasts of electric energy prices in medium and long-term electricity price forecasting. On the other hand, modeling approaches based on exponential smoothing are very popular in time series. These models are extensively used to accommodate multiple periodicities in time series data. In this method, the variable of interest is predicted as an exponentially weighted average of the sequenced past values. This method uses a smoothing factor known as the smoothing constant (lies between 0 and 1). It allocates the large weights to more recent observations and the weights decrease exponentially as the observations become more distant. Exponential smoothing approaches have been extensively used to forecast electricity load and prices [20]–[22]. For example, Cruz, *et al.* [22] used double seasonal exponential smoothing model to forecast hourly spot prices from the Spanish market. In their study, exponential smoothing performs slightly better than ARIMA, and both outperform the naïve method. Jónsson, *et al.* [23] used exponential smoothing techniques to predict the Nord Pool electricity market. All the models used in the paper were based on the well-known Holt–Winters model. The results suggested that exponential smoothing techniques are effective for predicting the day-ahead offering of wind power. Computational intelligence based techniques use nonparametric tools (e.g, artificial neural networks (ANNs), support vector machine (SVM), fuzzy logic, etc.) to model price processes. AI-based models are efficient in handling complexity and non-linearity, which makes them promising for short term predictions [24]. These methods generally map the relationship between input and output without exploring the underlying process. In the context of electricity prices forecasting, these methods have been used by some authors [25]–[33]. For example, a three-layered feedforward neural network, trained by the Levenberg-Marquardt algorithm, is used by Catalão, *et al.* [34] for forecasting next-week electricity prices from the electricity markets of mainland Spain and California. Based on their results, the neural network approach outperforms the ARIMA technique and the naive procedure in all considered weeks.

Even though a lot of models and techniques have been presented for electricity price forecasting, but none of them

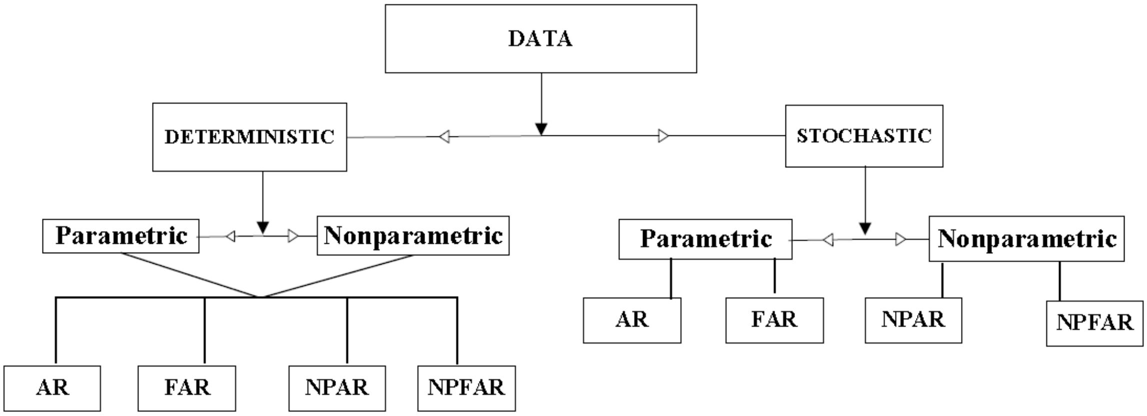

FIGURE 1. Flowchart of the proposed modeling framework.

provides accurate forecasts in a consistent manner. Therefore, a more efficient price forecasting framework is required since many market participants and system operators depend on it. This work considers the prices forecasting for Italian electricity market. Different from previous studies, this work investigates the role of parametric and nonparametric modeling approaches when considering models from both classes, univariate and multivariate. In the case of multivariate, this work considers a comparatively less explored modeling technique based on functional data analysis (FDA). In functional modeling, each daily price curve is considered as a single functional datum. Then, the daily electricity price profile (functional datum) is predicted in an autoregressive fashion using previous daily electricity price profiles. FDA is widely used in different scientific fields, but little explored in the context electricity market [35]–[37].

The rest of the article is organized as follows: Section 2 describes the methods and models used in this study. The description of the data and an application of the proposed methods and models to Italian electricity market is given in Section 3. The conclusion is given in Section 4.

II. GENERAL MODELING PROCEDURE

The main aim of this work is to forecast one-day-ahead electricity prices by using different forecasting techniques and models. As mentioned before, electricity price series is characterized by different complex dynamics. To accurately model these complex dynamics, first the price series is filtered out for the deterministic component and then the residual component is modeled by using different linear and non-linear models. After modeling both deterministic and stochastic components separately, the final forecast is obtained by combining the estimates of both components. To this end, let $\log(P_{d,l})$ be the logarithmic price series for d^{th} day (d $= 1, 2, \ldots, n$) and l^{th} load period ($l = 1, 2, \ldots, 24$). The model for the log price dynamics, $\log(P_{d,l})$, can be described as:

$$\log(P_{d,l}) = D_{d,l} + R_{d,l} \tag{1}$$

That is, the logarithmic price series, $\log(P_{d,l})$, is divided into two main components: a deterministic, $D_{d,l}$, component and a stochastic, $R_{d,l}$, component. The component $D_{d,l}$ consists of long-run dynamics, multiple periodicities (yearly and weekly cycles) and calendar effects. Mathematically, $D_{d,l}$ is modeled as:

$$D_{d,l} = T_{d,l} + Y_{d,l} + W_{d,l} + C_{d,l} \tag{2}$$

where $T_{d,l}$ refers to the long-run dynamic (trend), $Y_{d,l}$ and $W_{d,l}$ represents the yearly and weekly cycles, respectively, and $C_{d,l}$ denotes the calendar effect. Instead of including daily periodicity in 2, it is accounted by separately analyzing each load period. Finally, the stochastic component $R_{d,l}$ (residuals) which accounts for the short-run dynamics, is obtained as:

$$\begin{aligned} R_{d,l} &= \log(P_{d,l}) - \hat{D}_{d,l} \\ R_{d,l} &= \log(P_{d,l}) - (\hat{T}_{d,l} + \hat{Y}_{d,l} + \hat{W}_{d,l} + \hat{C}_{d,l}) \end{aligned} \tag{3}$$

and is modeled through different linear and non-linear models. Once both, deterministic and stochastic, components are estimated, final one-day-ahead prices forecast is obtained as:

$$\begin{aligned} \hat{P}_{d,l} &= \exp(\hat{T}_{d,l} + \hat{Y}_{d,l} + \hat{W}_{d,l} + \hat{C}_{d,l} + \hat{R}_{d,l}) \\ &= \exp(\hat{D}_{d,l} + \hat{R}_{d,l}) \end{aligned} \tag{4}$$

The flowchart of the proposed modeling framework is given in Figure 1.

A. MODELING DETERMINISTIC COMPONENT

This section explains the modeling and estimation of deterministic components $D_{d,l}$. For this purpose two approaches, parametric and nonparametric, are considered in this study. In the parametric approach, long-run component $T_{d,l}$ and the yearly component $Y_{d,l}$ are modeled through a linear regression model. For estimation of regression coefficients, Ordinary Least Square (OLS) technique is used. In nonparametric approach, the long-run component $T_{d,l}$ and the yearly component $Y_{d,l}$ are modeled by using smoothing splines where the series of $T_{d,l}$ and $Y_{d,l}$ are treated as a functional object. In both approaches, weekly cycles $W_{d,l}$ and bank holidays $C_{d,l}$ are described by dummy variables. For weekly cycles seven dummy variables are used i.e., $W_{d,l} = \sum_{j=1}^{7} \alpha_j I_{d,l}$ where $I_{d,l} = 1$ if d indicates the j^{th}

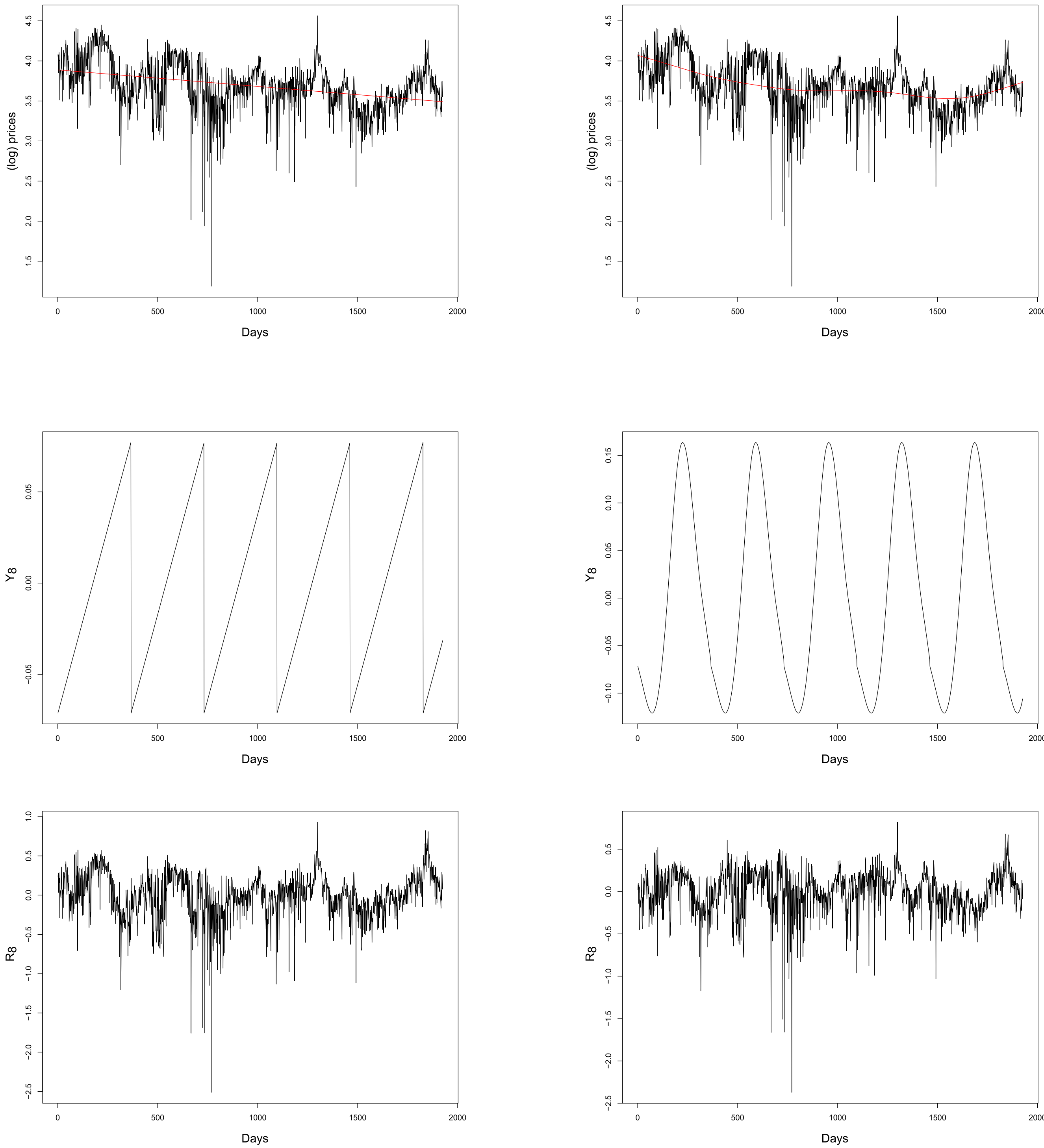

FIGURE 2. **IPEX price forecasting: (First row) logarithmic prices series for load period 8 with superimposed in red (right) parametrically estimated $T_{d,8}$ (left) nonparametrically estimated $T_{d,8}$, (Second row) (right) parametrically estimated $Y_{d,8}$ (left) nonparametrically estimated $Y_{d,8}$ (Third row) (right) $R_{d,8}$ from parametric model (left)$R_{d,8}$ from nonparametric model.**

day of the week and 0 otherwise. Similarly, two dummy variables are used for the bank holidays, i.e., $C_{d,l} = \sum_{j=1}^{2} \beta_j I_{d,l}$ where $I_{d,l} = 1$ if d represents a bank holiday and 0 otherwise. In the case of parametric estimation, all the components of $D_{d,l}$ are jointly estimated using the OLS technique whereas, back fitting algorithm is used in the case of nonparametric estimation to avoid curse of dimensionality.

An example of the estimated deterministic components are plotted in Figure 2. In the figure, the first row depicted the logarithmic prices for the load period 8 with superimposed in red parametrically (left) and nonparametrically (right) estimated long trend $T_{d,8}$. In the second row, parametrically estimated yearly component $Y_{d,8}$ (left) and its nonparametric counterpart is plotted in right panel. As the estimation of weekly component $W_{d,l}$ and bank holidays $C_{d,l}$ is similar in both cases, we prefer not to show their plots. The third row of Figure 2 refers to the stochastic component $R_{d,l}$ (residual) which is further modeled using parametric and nonparametric approaches.

B. MODELING STOCHASTIC COMPONENT

This section explains the modeling and estimation of stochastic (residuals) component $R_{d,l}$. To this end, two classes of models, i.e., univariate and multivariate (functional) models are used. Within each case, a parametric and a nonparametric approach is considered to specify the models. In the case of univariate models, parametric AutoRegressive (AR) and its nonlinear version, NonParametric AutoRegressive (NPAR) models are considered. On the other hand, a parametric Functional AutoRegressive (FAR) and a NonParametric Functional AutoRegressive (NPFAR) models are estimated. The details of each models are provided as under.

1) AutoRegressive MODELS

In univariate time series modeling, AutoRegressive (AR) models are perhaps the most widely used linear models. In an AR model, the variable of interest is regressed on its own p past values. The value p is called the order of an AR model, which indicates the number of past values required to predict the current value. Mathematically, an AR model of order p, AR(p), can be written as:

$$R_{d,l} = b + \beta_1 R_{d-1,l} + \beta_2 R_{d-2,l} + \ldots + \beta_p R_{d-p,l} + \epsilon_{d,l} \tag{5}$$

where $R_{d,l}$ is a stationary time series, b is a constant (intercept) term, β_j (j = 1, 2, ..., p) are parameters and $\epsilon_{d,l}$ is a white noise process, i.e., $\epsilon_{d,l} \sim WN(0, \sigma^2_\epsilon)$. The parameters of the model are estimated using the Maximum Likelihood (ML) method.

The choice of the lags used in the model is an important issue and is generally addressed by either looking at the ACF and PACF plots of the residual series or using some kind of information criteria, such as Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) etc. This work consider the inspection of the ACF and PACF plots of $R_{d,l}$ and selected a restricted AR(7) model with $\beta_3 = \cdots = \beta_6 = 0$.

2) NonParametric AutoRegressive MODELS

NonParametric AutoRegressive (NPAR) models are the extension of linear AR models. A Non-Parametric AutoRegressive (NPAR) model is built simply by permitting non-linearity in an AR model. NPAR models account the relationship between interest variable and its lagged variables without considering any specific parametric form. Mathematically, NPAR can be written as

$$R_{d,l} = g_1(R_{d-1,l}) + g_2(R_{d-2,l}) + \ldots + g_p(R_{d-p,l}) + \epsilon_{d,l} \tag{6}$$

where $g_k (k = 1, 2, \ldots, p)$ are smoothing functions representing the underling relations between the variable of interest and explanatory variables and $\epsilon_{d,l}$ is a random error term. The smoothing functions g_k can be described and estimated by using different techniques [38]. In this study, penalized cubic smoothing splines are used to define g_k. To prevent the issue of the curse of dimensionality, which is attributed to the exponential decay of data points within a smoothing window by increasing the dimension of regressors [39], an additive form is usually considered which assumes that the explanatory variables are uncorrelated. For the estimation purpose of g_k, back-fitting (iterative procedure) algorithm is used. Likewise, in the case of AR, the selection of the lags used in NPAR model is done by inspecting the ACF and PACF of $R_{d,l}$ and lag 1, 2, and 7 are used to fit the model.

3) FUNCTIONAL AutoRegressive MODELS

Functional data analysis (FDA) is a relatively new field where analysis is done by using information from curves or functions. In FDA, each datum is a single structured functional object. Due to this feature, FDA combines information on both, across and within sample units to make inference about the population. In addition, assuming the information as a functional object bypasses the problem of the numbers of variables and also allow to use additional information that may exist due to smoothness or derivatives contained in the functional structure. In FDA, the data are first converted to smooth functions known as basis function. In our case, the daily prices profile is considered as functional object (datum) and is converted to basis function by

$$\psi_d \equiv \psi_d(l) \equiv \delta(R_{d,1}, \ldots, R_{d,l}) = \sum_{k=1}^{K} a_k \gamma_k(l) \quad d = 1, \ldots, n$$

where a_k are constant parameters and $\gamma_k(j)$ are Fourier basis functions.

The functional version of the aforementioned classical linear autoregressive model is first introduced by [40], and is given by

$$\psi_d = \sum_{i=1}^{p} \int_{s \in (0,l)} \alpha_i(s) \psi_{d-i}(s) ds + \eta_d. \tag{7}$$

The kernels of the Hilbert-Schmidt operators, $\alpha_i(s)$ are the functional parameters of the model, and η_d is the functional error. This model is referred to as a functional autoregressive model of order p ($FAR(p)$) and it allows to predict daily prices curves conditionally to past observed curves. This model is estimated using the methodology described in [41] that deals with the covariance and cross-covariance operators estimation of a Hilbert space (H)-valued autoregressive process. Furthermore, FAR(1) model is used for modeling and forecasting purposes.

TABLE 1. IPEX price forecasting: Mean absolute percentage error (MAPE), Mean absolute error (MAE) and Person correlation coefficient (r) for one-day-ahead out-of-sample forecast using different approaches for deterministic component *D* and different models for stochastic component *R*.

Estimation of D	Estimation of R	MAE	MAPE	r
Parametric	AR	5.6544	9.9237	0.8480
	NPAR	5.4888	**9.7375**	0.8597
	FAR	**5.4806**	9.7644	**0.8642**
	NPFAR	5.8287	10.2841	0.8515
Nonparametric	AR	5.8638	10.6170	0.8402
	NPAR	5.6630	10.2687	0.8531
	FAR	**5.6300**	**10.1974**	**0.8596**
	NPFAR	6.1643	11.2923	0.8375
Naïve1		8.3821	15.2488	0.6937
Naïve2		6.8414	12.5415	0.8095

4) NonParametric FUNCTIONAL AutoRegressive MODELS

The nonlinear version of the above stated functional model is called the NonParametric Functional AutoRegressive model (NPFAR). In this case, we do not assume any parametric form of the relationship between response and predictors, thus allowing for any possible nonlinear relationship. Mathematically, it is defined as

$$\psi_d = \frac{\sum_{t=p+1}^{n} \psi_d K_b(\xi(\psi_{d-p}, \tilde{\psi}))}{\sum_{t=p+1}^{n} K_b(\xi(\psi_{t-p}, \tilde{\psi}))} \tag{8}$$

where $\psi_d(.)$ is a functional random variable valued in some semi−metric space (E, ξ), $\tilde{\psi}$ is a given element of E, $K(\cdot)$ is a probability density function (kernel) with $K_b(u) = K(u/b)$ and b represents the smoothing parameter (bandwidth). Estimation procedure and more details about this model can be found in [42]. Finally, likewise FAR, this work consider NPFAR(1) model for modeling and forecasting purposes.

C. THE BENCHMARK MODELS

We conclude this section by introducing two benchmark models that belong to the similar-day technique. The first benchmark model will be called Naïve1 method proposed by [17]. In this method, electricity price forecast of hour l on Saturday, Sunday, and Monday are determined by using prices on l hour a week ago on the same day, i.e., $\hat{p}_{d,l} = p_{d-7,l}$. While for Tuesday, Wednesday, Thursday and Friday, the electricity price forecast for hour l is set equal to the price for the same hour on the previous day, i.e., $\hat{p}_{d,l} = p_{d-1,l}$.

The second benchmark model, Naïve2, also belong to the similar-day technique. This method proceeds as follows. To forecast, for example, Monday, we select the day before Monday, i.e. Sunday. We then select all the previous Sundays in the data and compare every Sunday independently with the current Sunday price profile. The difference between any previous Sunday and current Sunday is summarized using MAE (Other statistics can also be used, e.g. MAPE etc., however we found MAE to be more useful in forecasting). In this way we obtain a vector of MAE values for each comparison. We then find the Sunday that produces the smallest MAE value when compared to the current Sunday. Once the Sunday is identified, we use the next day to the identified Sunday, i.e.,Monday and use it as the forecast for the Monday we are interested. We do the same procedure for all days of the week.

III. OUT-OF-SAMPLE PRICE FORECASTING

In this work, the electricity prices data from the Italian Electricity Market (IPEX) are considered. The data set ranges from January 1st 2012 to December 31th 2017 (52608 observations, covering 2192 days). Each day of the data set comprises of 24 observations, where each observation corresponds to a load period. For modeling and forecasting purposes, the data set is further divided into two sets: January 1st 2012 to December 31th 2016 (43848 observations, covering 1827 days) for identification and estimation of the models, and January 1st 2017 to December 31th 2017 (8760 observations, covering 365 days) for evaluating one-day-ahead out-of-sample forecasting accuracy of the models. For each load period, all the models are separately estimated for deterministic component, and when considering univariate modeling of stochastic component. For the whole 2017 year, one-day-ahead forecasts are obtained by using the expanding window method. For comparing the forecasting accuracy of the models, three different types of descriptive statistics are used. These includes, Mean Absolute Error (MAE), Mean Absolute Percentage Errors (MAPE), and Pearson correlation coefficient (r). Mathematically, they are defined as:

$$\text{MAE} = \text{Mean}\left(\mid P_{d,l} - \hat{P}_{d,l} \mid \right)$$

$$\text{MAPE} = \text{Mean}\left(\frac{\mid P_{d,l} - \hat{P}_{d,l} \mid}{P_{d,l}} \right) \cdot 100$$

$$r = \text{corr}(P_{d,l}, \hat{P}_{d,l})$$

where $P_{d,l}$ and $\hat{P}_{d,l}$ are the observed and forecasted prices for d^{th} day (d $=$ $1, 2, \ldots, 365$) and l^{th} load period (l $= 1, 2, \ldots, 24$), respectively.

The results for one-day-ahead out-of-sample forecasting for the complete year 2017 are listed in Table 1. The first column of the table refers to the estimation of the deterministic component either through, parametric or nonparametric estimation. The second column listed the models used for the estimation of the stochastic component. By combining models from both, deterministic and stochastic, components

TABLE 2. 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).

Estimation of D		Parametric				Nonparametric					
		AR	NPAR	FAR	NPFAR	AR	NPAR	FAR	NPFAR	Naïve1	Naïve2
Parametric	AR	-	0.99	0.96	0.32	0.27	0.36	0.44	0.04	< 0.01	< 0.01
	NPAR	0.01	-	0.55	0.01	0.01	0.02	0.05	0.01	< 0.01	< 0.01
	FAR	0.04	0.45	-	0.01	0.02	0.02	0.04	0.01	< 0.01	< 0.01
	NPFAR	0.68	0.99	0.99	-	0.38	0.58	0.76	0.11	< 0.01	< 0.01
Nonparametric	AR	0.73	0.99	0.98	0.62	-	0.99	0.96	0.24	< 0.01	< 0.01
	NPAR	0.64	0.98	0.98	0.42	0.01	-	0.79	0.01	< 0.01	< 0.01
	FAR	0.56	0.95	0.96	0.24	0.04	0.21	-	0.01	< 0.01	< 0.01
	NPFAR	0.96	0.99	0.99	0.89	0.76	0.99	0.99	-	< 0.01	0.14
Naïve1		> 0.99	> 0.99	> 0.99	> 0.99	> 0.99	> 0.99	> 0.99	> 0.99	-	> 0.99
Naïve2		> 0.99	> 0.99	> 0.99	> 0.99	> 0.99	> 0.99	> 0.99	0.86	< 0.01	-

TABLE 3. IPEX price forecasting: Day-specific mean absolute percentage error (DS-MAPE) for one-day-ahead out-of-sample forecast using different approaches for deterministic component *D* and different models for stochastic component *R*.

Estimation of D	Parametric				Nonparametric			
Weekday	AR	NPAR	FAR	NPFAR	AR	NPAR	FAR	NPFAR
Monday	10.94	11.02	11.54	11.74	12.15	11.96	12.24	13.02
Tuesday	10.11	9.83	9.34	10.29	10.65	10.41	10.34	11.23
Wednesday	10.12	9.34	9.30	10.16	10.94	10.17	9.33	10.65
Thursday	9.47	8.78	8.39	8.75	10.07	9.44	8.86	10.10
Friday	9.78	9.02	9.12	8.93	10.04	9.10	9.11	9.64
Saturday	9.15	10.03	9.18	10.77	9.91	10.29	10.16	11.53
Sunday	9.98	10.22	11.59	11.44	10.57	10.56	11.39	12.93

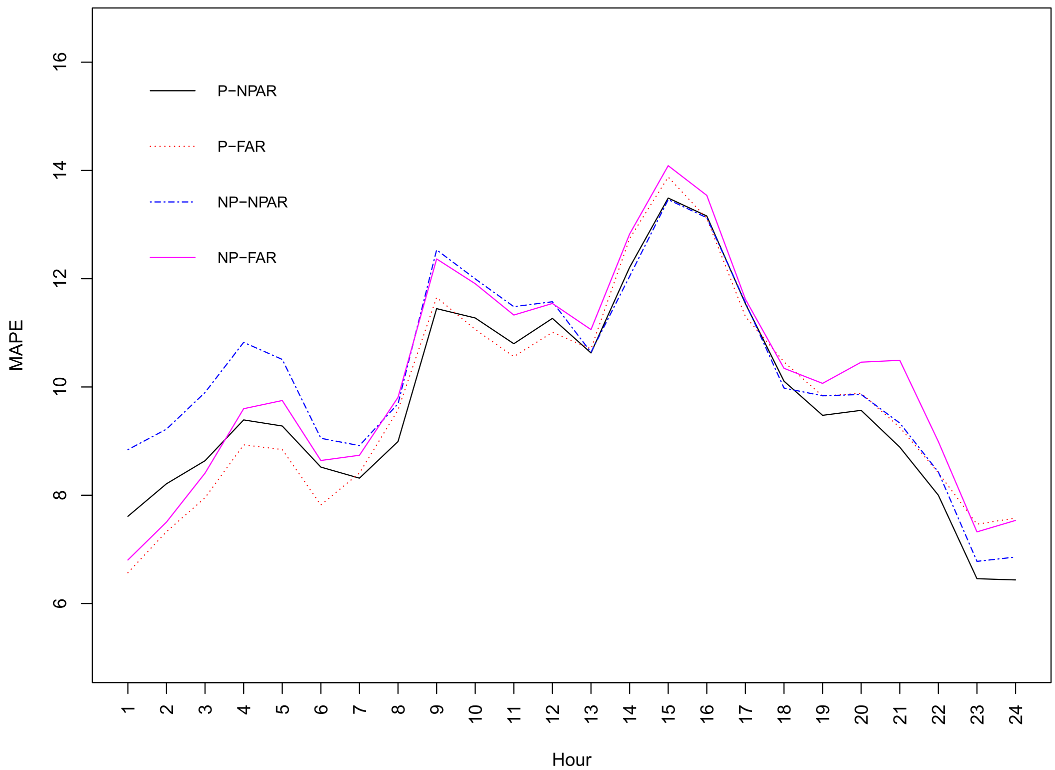

FIGURE 3. IPEX price forecasting: Hourly MAPE values for NPAR and FAR models with both, parametric (P) and nonparametric (NP) estimation of deterministic component *D*.

give eight models to compare. The results obtained from benchmark models are also listed in the last two rows of Table 1.

The results suggest that when the estimation of deterministic component is done through parametric approach, the models FAR and NPAR perform relatively better. The MAE

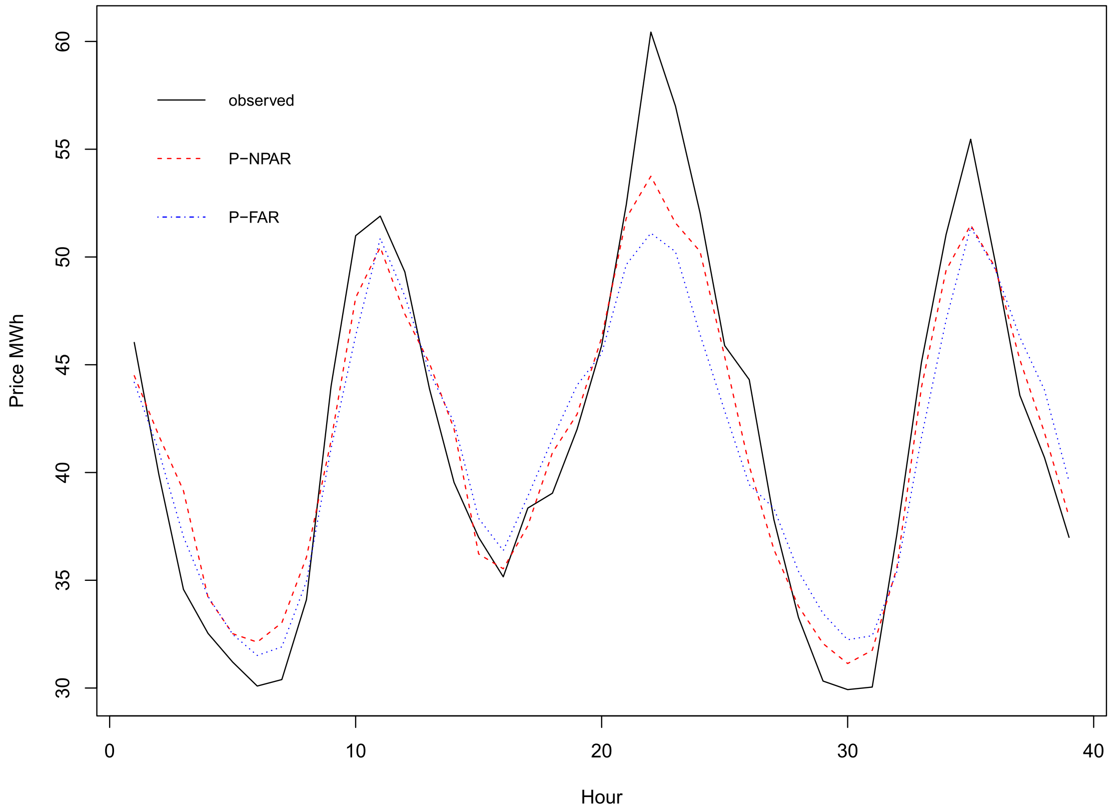

FIGURE 4. IPEX price forecasting: Observed prices (solid) with forecasted prices by NPAR (red dashed) and FAR (blue dotted) when deterministic component *D* is estimated through parametric approach.

and Pearson correlation coefficient, r, values are slightly better for FAR while NPAR produced comparatively small MAPE value. On the other hand, when the deterministic component is estimated nonparametrically, the FAR produces better MAE, MAPE and r values compared to the rest. Note that, the overall forecasting is better when the estimation of deterministic component is done with parametric approach. Secondly, as FAR produced slightly better results in both cases, it may indicate that parametric functional modeling is relatively better than the univariate. Furthermore, all the autoregressive models outperform both the naïve benchmarks. It is worth mentioning that Naïve2 performs much better compared to Naïve1.

To access the significance of the differences among the results listed in Table 1, Diebold and Mariano (DM) [43] test is performed and the results are listed in Table 2. The null hypothesis of the DM test corresponds to equal forecast accuracy while the alternative hypothesis states that model in row is more accurate than in the column. The results in the table confirms the superiority of NPAR and FAR models compared to the rest, especially when the deterministic component is estimated through parametric approach. Further, the forecasting accuracy of FAR is no longer superior than NPAR.

To see the performance of models on different days of the week, day-specific MAPE values are listed in Table 3. From this table one can see that MAPE values are comparatively higher for Monday, Saturday, and Sunday than the remaining week days. As the weekend comprised of Saturday and Sunday, these days generally have different pattern of electricity prices compared to the working days, thus forecasting error is high. On the other hand, Monday is followed by Saturday and Sunday, thus models perform relatively poor to capture Monday pattern. As NPAR and FAR models perform relatively better compared to AR and NPFAR, the hourly MAPE (H-MAPE) values for these two models are depicted in Figure 3, when the deterministic component *D* is estimated by using parametric and nonparametric approaches. From the figure one can see that H-MAPE is relatively lower in the initial and final hours of the day while it is higher in the middle load periods of the day. The H-MAPE is relatively higher for NP-NPAR till mid-day whereas it starts decreasing for the remaining hours. The forecasting superiority of P-NPAR and P-FAR is evident from the graph as they show minimum MAPE values on most load periods. The forecasted prices time series for the two best models along with the observed prices are plotted in Figure 4. From the figure it is evident that both the models are following the observed series quite well.

Finally, the forecasting performance of our models is highly competitive compared to the results reported in the literature. Note that it is generally difficult to compare results from other research work as different authors used different forecasting horizon, forecasting period and indicators.

For example, [44] reported a root mean square error (RMSE) of 11.58 obtained with an ARX-EGARCH model, whereas we obtained 8.43 (RMSE are not reported here) with the P-FAR. [45] obtained a RMSE of 16.72 and 15.79 using ARMA and GARCH models respectively that is considerably higher than our value of 8.43. The work of [46] resulted in MAE of 8.58 which is significantly higher than our MAE value of 5.48.

IV. CONCLUSION

The main aim of this work is to model and forecast electricity price time series. To this end, a components estimation method is used where the price time series is divided into two major components: deterministic and stochastic. The deterministic component consists of long-run dynamics, multiple periodicities (yearly and weekly cycles) and calendar effects whereas the stochastic component accounts for the short-run dynamics of the process. Deterministic as well as stochastic components are modeled through parametric and nonparametric approaches. For the estimation of stochastic component, two univariate models, i.e., AutoRegressive (AR) and Nonparametric AutoRegressive (NPAR), and two multivariate models, i.e., Functional AutoRegressive (FAR), and Nonparametric Functional AutoRegressive (NPFAR), are used. After modeling and forecasting both deterministic and stochastic components separately, the final forecasts are obtained by combining the estimates of both components. The electricity price data, ranges from January 1st 2012 to December 31 2017, are used from the Italian electricity market (IPEX). The results indicate that the component estimation approach is efficient in forecasting electricity prices series. The parametric estimation of deterministic component leads to better forecasting results. Finally, the functional model FAR produced better results compared to the rest models while the univariate model NPAR was a good competitor. Since, the results of this work are based on only IPEX, empirical analysis conducted on other electricity markets are recommended for the future studies. In addition, one may investigate the possibility to include exogenous variables, like temperature, oil prices, forecasted electricity demand, etc. to the model in the future work.

REFERENCES

[1] R. J. Serrallés, "Electric energy restructuring in the European union: Integration, subsidiarity and the challenge of harmonization," *Energy Policy*, vol. 34, no. 16, pp. 2542–2551, Nov. 2006.

[2] L. Hu, G. Taylor, H.-B. Wan, and M. Irving, "A review of short-term electricity price forecasting techniques in deregulated electricity markets," in *Proc. 44th Int. Universities Power Eng. Conf. (UPEC)*, Sep. 2009, pp. 1–5.

[3] G. Li, C.-C. Liu, C. Mattson, and J. Lawarree, "Day-ahead electricity price forecasting in a grid environment," *IEEE Trans. Power Syst.*, vol. 22, no. 1, pp. 266–274, Feb. 2007.

[4] R. Weron, "Heavy tails and electricity prices," Presented at the Deutsche Bundesbank's Annu. Fall Conf. (Eltville), Nov. 2005.

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

[6] R. Weron, *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*, vol. 403. Hoboken, NJ, USA: Wiley, 2007.

[7] P. Mandal, T. Senjyu, N. Urasaki, T. Funabashi, and A. K. Srivastava, "Short-term price forecasting for competitive electricity market," in *Proc. 38th North Amer. Power Symp.*, Sep. 2006, pp. 137–141.

[8] S. S. Torbaghan, A. Motamedi, H. Zareipour, and L. A. Tuan, "Medium-term electricity price forecasting," in *Proc. North Amer. Power Symp. (NAPS)*, Sep. 2012, pp. 1–8.

[9] M. Cerjan, A. Petričić, and M. Delimar, "HIRA model for short-term electricity price forecasting," *Energies*, vol. 12, no. 3, p. 568, Feb. 2019.

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

[11] T. Jakaša, I. Andročec, and P. Sprčić, "Electricity price forecasting—ARIMA model approach," in *Proc. 8th Int. Conf. Eur. Energy Market (EEM)*, May 2011, pp. 222–225.

[12] A. J. Conejo, M. A. Plazas, R. Espinola, and A. B. Molina, "Day-ahead electricity price forecasting using the wavelet transform and ARIMA models," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 1035–1042, May 2005.

[13] Z. Tan, J. Zhang, J. Wang, and J. Xu, "Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models," *Appl. Energy*, vol. 87, no. 11, pp. 3606–3610, Nov. 2010.

[14] K. Maciejowska and R. Weron, "Forecasting of daily electricity prices with factor models: Utilizing intra-day and inter-zone relationships," *Comput. Statist.*, vol. 30, no. 3, pp. 805–819, Sep. 2015.

[15] H. Liu and J. Shi, "Applying ARMA–GARCH approaches to forecasting short-term electricity prices," *Energy Econ.*, vol. 37, pp. 152–166, May 2013.

[16] F. Ziel, "Forecasting electricity spot prices using lasso: On capturing the autoregressive intraday structure," *IEEE Trans. Power Syst.*, vol. 31, no. 6, pp. 4977–4987, Nov. 2016.

[17] F. J. Nogales, J. Contreras, A. J. Conejo, and R. Espinola, "Forecasting next-day electricity prices by time series models," *IEEE Trans. Power Syst.*, vol. 17, no. 2, pp. 342–348, May 2002.

[18] Â. P. Ferreira, J. G. Ramos, and P. O. Fernandes, "A linear regression pattern for electricity price forecasting in the iberian electricity market," *Revista Facultad de Ingeniería Universidad de Antioquia*, vol. 93, pp. 117–127, Aug. 2019.

[19] S. Bordignon, D. W. Bunn, F. Lisi, and F. Nan, "Combining day-ahead forecasts for british electricity prices," *Energy Econ.*, vol. 35, pp. 88–103, Jan. 2013.

[20] T. Jonsson, P. Pinson, H. A. Nielsen, H. Madsen, and T. S. Nielsen, "Forecasting electricity spot prices accounting for wind power predictions," *IEEE Trans. Sustain. Energy*, vol. 4, no. 1, pp. 210–218, Jan. 2013.

[21] R. Beigaite, T. Krilavicius, and K. L. Man, "Electricity price forecasting for nord pool data," in *Proc. Int. Conf. Platform Technol. Service (PlatCon)*, Jan. 2018, pp. 1–6.

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

[23] T. Jónsson, P. Pinson, H. Nielsen, and H. Madsen, "Exponential smoothing approaches for prediction in real-time electricity markets," *Energies*, vol. 7, no. 6, pp. 3710–3732, Jun. 2014.

[24] R. Weron and A. Misiorek, "Short-term electricity price forecasting with time series models: A review and evaluation," Hugo Steinhaus Center, Wroclaw Univ. Technol., Wroclaw, Poland, Tech. Rep., HSC Res. Rep. HSC/06/01, 2006.

[25] M. Hayati and Y. Shirvany, *Artificial Neural Network Approach for Short Term Load Forecasting for Illam Region*, vol. 28. Paris, France: World Academy of Science, Engineering and Technology, 2007, pp. 280–284.

[26] J. Zhang and C. Cheng, "Day-ahead electricity price forecasting using artificial intelligence," in *Proc. IEEE Canada Electr. Power Conf.*, Oct. 2008, pp. 1–5.

[27] B. Neupane, K. S. Perera, Z. Aung, and W. L. Woon, "Artificial neural network-based electricity price forecasting for smart grid deployment," in *Proc. Int. Conf. Comput. Syst. Ind. Informat.*, Dec. 2012, pp. 1–6.

[28] R. Burbidge and B. Buxton, "An introduction to support vector machines for data mining," in *Keynote Papers, Young OR12*. Nottingham, U.K.: Univ. of Nottingham, Operational Research Society, 2001, pp. 3–15.

[29] V. Jakkula, *Tutorial on Support Vector Machine (SVM)*, vol. 37. Washington, DC, USA: School of EECS, 2006.

[30] X. Yan and N. A. Chowdhury, "Midterm electricity market clearing price forecasting using two-stage multiple support vector machine," *J. Energy*, vol. 2015, pp. 1–11, Jan. 2015.

[31] C. P. Rodriguez and G. J. Anders, "Energy price forecasting in the ontario competitive power system market," *IEEE Trans. Power Syst.*, vol. 19, no. 1, pp. 366–374, Feb. 2004.

[32] J. P. S. Catalao, H. M. I. Pousinho, and V. M. F. Mendes, "Hybrid Wavelet-PSO-ANFIS approach for short-term electricity prices forecasting," *IEEE Trans. Power Syst.*, vol. 26, no. 1, pp. 137–144, Feb. 2011.
[33] A. Alshejari and V. S. Kodogiannis, "Electricity price forecasting using asymmetric fuzzy neural network systems," in *Proc. IEEE Int. Conf. Fuzzy Syst. (FUZZ-IEEE)*, Jul. 2017, pp. 1–6.
[34] J. P. S. Catalão, S. J. P. S. Mariano, V. M. F. Mendes, and L. A. F. M. Ferreira, "Short-term electricity prices forecasting in a competitive market: A neural network approach," *Electr. Power Syst. Res.*, vol. 77, no. 10, pp. 1297–1304, Aug. 2007.
[35] I. Shah and F. Lisi, "Day-ahead electricity demand forecasting with non-parametric functional models," in *Proc. 12th Int. Conf. Eur. Energy Market (EEM)*, May 2015, pp. 1–5.
[36] D. Liebl, "Modeling and forecasting electricity spot prices: A functional data perspective," *Ann. Appl. Statist.*, vol. 7, no. 3, pp. 1562–1592, Sep. 2013.
[37] I. Shah and F. Lisi, "Forecasting of electricity price through a functional prediction of sale and purchase curves," *J. Forecasting*, vol. 39, no. 2, pp. 242–259, Mar. 2020.
[38] T. J. Hastie and R. J. Tibshirani, *Generalized Additive Models*. Boca Raton, FL, USA: CRC Press, 1990.
[39] L. Wasserman, *All of Nonparametric Statistics*. New York, NY, USA: Springer-Verlag, 2006.
[40] J. O. Ramsay and C. Dalzell, "Some tools for functional data analysis," *J. Roy. Stat. Soc., Ser. B, Methodol.*, vol. 53, pp. 539–572, Jul. 1991.
[41] D. Bosq, *Linear Processes in Function Spaces: Theory and Applications*, vol. 149. New York, NY, USA: Springer-Verlag, 2000.
[42] F. Ferraty, I. Van Keilegom, and P. Vieu, "Regression when both response and predictor are functions," *J. Multivariate Anal.*, vol. 109, pp. 10–28, Aug. 2012.
[43] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *J. Bus. Econ. Stat.*, vol. 13, no. 3, pp. 253–263, 1995.
[44] 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.
[45] 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.
[46] A. Petrella and S. Sapio, "No PUN intended: A time series analysis of the Italian day-ahead electricity prices," Florence School Regulation, Loyola de Palacio Programme Energy Policy, EUI RSCAS, Tech. Rep., Mar. 2010. [Online]. Available: http://hdl.handle.net/1814/13099

ISMAIL SHAH received the master's degree from Lund University, Sweden, and the Ph.D. degree from the University of Padova, Italy. He is currently an Assistant Professor with the Department of Statistics, Quaid-i-Azam University (QAU), Islamabad, Pakistan. His research interests include functional data analysis, time series analysis, energy economics, and applied and industrial statistics.

HAJRA BIBI received the M.Phil. degree in statistics from Quaid-i-Azam University (QAU), Islamabad, Pakistan. Her research interests include time series analysis, forecasting electricity market variables, and applied statistics.

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

LICHEN WANG received the Ph.D. degree in crop genetics and breeding from Huazhong Agricultural University, Wuhan, China, in 2017. She is currently a Lecturer with the College of Life Science, Linyi University, China. Her current research interests include high throughput sequence data analysis, transcription factor functional characterization in plants, and IncRNAs and miRNAs functional analysis in cotton fiber initiation and elongation.

ZHEN YUE received the Ph.D. degree from the College of Agriculture and Biotechnology, China Agricultural University, Beijing, China, in 2016. She is currently a Lecturer with the College of Life Science, Linyi University, Linyi, China. Her research interests include biological control, insect physiology, biochemistry, molecular biology, and quantitative analysis.