Renewable and Sustainable Energy Reviews 94 (2018) 251–266

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

Probabilistic mid- and long-term electricity price forecasting

Florian Ziel[a,*], Rick Steinert[b]

[a] Universität Duisburg-Essen, Berliner Platz 6-8, 45127 Essen, Germany
[b] Europa-Universität Viadrina, Große Scharrnstraße 59, 15230 Frankfurt, Oder, Germany

ARTICLE INFO

Keywords:
Electricity prices
Probabilistic forecasting
Supply and demand
Long-term
Negative prices
Renewable energy

ABSTRACT

The liberalization of electricity markets and the development of renewable energy sources has led to new challenges for decision makers. These challenges are accompanied by an increasing uncertainty about future electricity price movements. The increasing amount of papers, which aim to model and predict electricity prices for a short period of time provided new opportunities for market participants. However, the electricity price literature seem to be very scarce on the issue of medium- to long-term price forecasting, which is mandatory for investment and political decisions. Our paper closes this gap by introducing a new approach to simulate electricity prices with hourly resolution for several months up to three years. Considering the uncertainty of future events we are able to provide probabilistic forecasts which are able to detect probabilities for price spikes even in the long-run. As market we decided to use the EPEX day-ahead electricity market for Germany and Austria. Our model extends the X-Model which mainly utilizes the sale and purchase curve for electricity day-ahead auctions. By applying our procedure we are able to give probabilities for the due to the EEG practical relevant event of six consecutive hours of negative prices. We find that using the supply and demand curve based model in the long-run yields realistic patterns for the time series of electricity prices and leads to promising results considering common error measures.

1. Introduction and motivation

The past decades in electricity price research were characterized by the rapid liberalization process of several electricity markets across the world and the increasing development of renewable energy. Either voluntarily or by regulation, many institutions in the field of electricity contributed to a continuously improving transparency and quality of mostly freely available information on electricity prices and related time series. This in turn has helped researchers and practitioners to understand the mechanics of the price formation and lead to a large amount of papers which focus on electricity price forecasting. According to Weron, there was only a negligible amount of papers published before the year 2000, whereas in 2005 and 2006 the amount of papers reached their first peak point followed up by its hitherto maximum in 2009 [1].

Research in electricity price forecasting originates from many different fields of science, e.g. engineering or statistics, which led to a manifold structure of different approaches. However, most of these approaches have in common that they focus on forecasting electricity prices in the short-term, specifically up to one day ahead with an hourly resolution (see e.g. [2] or [1] for a literature review on electricity price forecasting). In contrast to this, electricity price forecasting methods

Abbreviation: ACE, average coverage error; AIC, Akaike information criterion; AMAPE, adapted MAPE; ANEM, Australian National Electricity Market; ANN, artificial NN; ANOVA, analysis of variance; AWPI, average width of PIs; ARMAX, autoregressive moving average model with exogenous inputs; ARX, autoregressive model with exogenous inputs; BNetzA, German Federal Network Agency; BS, Brier Score; CRPS, continuous ranked probability score; CT, Christoffersen test; CWC, coverage width-based criterion; DC, direct current; DWD, German Meteorological Office; ECP, empirical coverage probability; ECR, evaluation criterion of resolution; EEG, German Renewable Energy Sources Act; ELM, extreme learning machine; ENTSO-E, European Network of Transmission System Operators; EPEX, European Power Exchange; EUPHEMIA, Pan-European Hybrid Electricity Market Integration Algorithm; GAMLSS, generalized additive models for location, scale and shape; GARCH, generalized autoregressive conditional heteroskedasticity; GDP, gross domestic product; GME, Gestore dei Mercati Energetici; k-NN, k nearest neighbor; Lasso, least absolute shrinkage and selection operator; LS, logarithmic scores; LSSVM, least-squares SVM; MAE, mean absolute error; MAPE, mean absolute percentage error; MLP, multi-layer perceptron; MPIW, mean prediction interval width; MSE, mean squared error; MSPE, mean squared percentage error; NN, neural network; NLPD, average negative log predictive density; NMPIW, normalised MPIW; OLS, ordinary least squares; OMIE, OMI-Polo Español; PBS, pinball score/loss; PCA, principal component analysis; PCR, Price Coupling of Regions; PI, prediction interval; PICP, PI coverage probability; PINAW, PI normalised average width; PITS, probability integral transform scores; PJM, Pennsylvania-New Jersey-Maryland Interconnection; RBF, radial basis function; RE, reliability evaluation criterion; RMSE, root MSE; SARIMAX, seasonal autoregressive integrated moving average with exogenous inputs; SE, sharpness criterion/score; SVM, support vector machine; UK, United Kingdom; US, United States; VEC, vector error correction model; WNN, weighted nearest neighbor; WS, Winkler score

* Corresponding author.
E-mail address: florian.ziel@uni-due.de (F. Ziel).

https://doi.org/10.1016/j.rser.2018.05.038
Received 31 March 2017; Received in revised form 15 May 2018; Accepted 16 May 2018

which consider a longer period of time are rare [3]. A large proportion of research for that time horizon originates from fundamental models, which capture the dynamics of the system, e.g. the estimated cost functions of the market participants [4,5]. These model types often lack to use realistic time series of prices and related data and therefore cannot provide a realistic hourly resolution of price predictions, which is typically the case in day-ahead markets.

Nevertheless, there are some models which are able to capture the hourly behavior of electricity price and provide mid- to long-term forecasts. Even though the literature is not consentaneous on this issue, we refer to the time horizon of one month to one year as mid-term and to the time horizon of more than one year as long-term. The model of [6] for instance consists of a hybrid approach using fundamental and econometric, e.g. autoregressive, modeling techniques. They are able to utilize the hourly day-ahead electricity price series of Spain to forecast the whole year 2005. Yan and Chowdhurry were able to use data mining techniques, e.g. support vector machines (SVM), to study the PJM market in 2013 and 2015. In 2013 they show by a forecasting study that combining a least squares support vector machine with an ARMAX model yields promising results when the hourly forecast of one month is considered. For their setup they use training data of one year, e.g. 2009, to forecast the month of July 2010 [3]. Applying a two-stage SVM in 2015 they extend their model to be able to capture severe price peaks, which they describe as extremely difficult to model in a mid-term forecasting setting [7]. As in their previous paper, they forecast one month with hourly resolution.

Another important limitation of most electricity price forecasting models is their focus on specific moments of the distribution, particularly the mean and the variance. Also machine learning techniques commonly concentrate on point forecasts - which is a comparable counterpart to forecasting the mean in an econometric setting. Even though it seems most important at first glance to get point or mean forecasts for the electricity price, it is often the uncertainty of prices which has the highest impact for market participants. Achieving a precise point forecast may provide a solid basis for flexible investment decisions, but cannot account for the likelihood of possible extreme events, which can have tremendous consequences for the business as a whole. However, some researchers tackle this issue by analyzing and modeling the variance as well. But the concept of variance alone is not enough to quantify uncertainty in the case of electricity prices, as they usually tend to have non-symmetric heavy-tailed distributions which also vary over time. Hence, a possible solution for this issue can be to model the whole time-dependent distribution function of prices. This field of research was considered in electricity price forecasting especially in recent years. It can be summarized under the discipline of probabilistic forecasting. One of the early papers covering probabilistic forecasts in terms of interval forecasts originated in 2006 by [8]. They utilize well-known point forecasting models like ARX with GARCH components to construct interval forecasts for the hourly electricity price of the California Power Exchange. Later contributions, which explicitly focus on probabilistic forecasting emerged from econometric as well as machine learning approaches. For instance, Kou et al. were able to achieve day-ahead probabilistic forecasts for several electricity markets by combining machine learning with a variational heteroscedastic Gaussian process [9]. A common approach during the recent years for econometric probabilistic forecasting was constructing prediction intervals by quantile regression. A basic introduction for this topic can be found in [10], among others. Extensions include for instance the Factor Quantile Regression Averaging of [11] or lasso-based approaches as done by [12].

A recent review done by [13] focused on the raising awareness of probabilistic forecasting in electricity price forecasting. They support their argument of an increased necessity of these methods by quantifying the development of published articles from the year 2003, where the first related article was published by [14] and the year 2016. Given their numbers they show that probabilistic electricity price forecasting gained a tremendous increase in 2016, when the amount of published papers almost quadrupled from 3 to 11 in the year 2016 by the time of their study.

Another important direction of electricity price models originates from fundamental or structural electricity price models. For these models the electricity price is considered as an equilibrium of supply and demand (see e.g. [15–17]). Here the major price drivers are the fundamental inputs like the load and the merit order curve and especially the marginal cost of the available power plant portfolio. These models are popular for long term forecasting of electricity prices as impacts of regulative changes, for example the closing of a certain power plant or a newly installed wind farm, can be easily drawn. However, these models usually only model the equilibrium price, e.g. the mean electricity price at a certain time point, but not the full underlying distribution. Even though new approaches in this direction like [18] try to overcome this problem particularly by providing a representative fundamental market situation, the general problem that no temporal dependency information is used remains unsolved.

Many researchers who conduct a review study seem to mainly focus on the difference in models and try to compare them by presenting overviews or their popularity over time ([2,13,1]; among others). This focus resembles a point of view where the method is of utmost importance and not necessarily the purpose of modeling. However, from a practical standpoint it could be argued that the way how electricity prices are modeled is not as much important as the goal the modeling strategy actually pursues. For instance, if an electricity company is interested in building a new power plant, they will mainly be interested in long term electricity price forecasts over the whole lifetime of the plant, rather than to focus on a short-term horizon. Comparing different model strategies however may prove not to be too useful in this situation as it requires deep knowledge about properties and limitations of these models. Therefore we decided to present a brief review of the literature with the focus of the actual purpose of modeling rather than the model itself. This review part can therefore be considered as an update to e.g. [13] with taking a new point of view on the issue. Hence, we will continue with a review of the recent and practical relevant topics of forecasting horizon and probabilistic forecasting.

For our study we divided these two categories in the following subcategories. Forecasting horizon was divided into mid-term and long-term forecasting as well as a category which accumulates over all horizons. This last category was chosen to easily determine the relative amount of mid- and long-term forecasting compared to all price forecasting horizons. Probabilistic forecasting is a binary category which acknowledges a paper as probabilistic forecasting whenever the full density of prices is forecasted. This specifically excludes papers which only forecast mean or variance. To analyze the amount of papers published in this field, we used Scopus as it is not only a well-known and reliable source for papers but also has an user-friendly interface for refined queries. To achieve rigour we geared the applied keywords towards the study of [13] with minor changes. These keywords were combined by logical links to create the database-syntax specific queries for Scopus with the requirements for our categories. Hence, we needed to conduct six different queries, one for all electricity price forecasting papers,[1] one for mid-term forecasting,[2] another one for long-term

[1] (TITLE((((("electric*" OR "energy market" OR "power price" OR "power market" OR "power system" OR pool OR "market clearing" OR "energy clearing") AND (price OR prices OR pricing)) OR lmp OR "locational marginal price") AND (forecast OR forecasts OR forecasting OR prediction OR predicting OR predictability OR "predictive densit*")) OR ("price forecasting" AND "smart grid*")) OR TITLE-ABS("electricity price forecasting" OR "forecasting electricity price" OR "day-ahead price forecasting" OR "day-ahead mar* price forecasting" OR (gefcom2014 AND price) OR (("electricity market" OR "electric energy market") AND "price forecasting") OR ("electricity price" AND ("prediction interval" OR "interval forecast" OR "density forecast" OR "probabilistic forecast"))) AND NOT TITLE ("unit commitment")) AND (EXCLUDE(AU-ID,"[No Author ID found]" undefined)).

[2] Query of footnote 1 combined with: AND (TITLE("mid-term") OR TITLE("mid term")