Energy Economics 33 (2011) 1216–1226

Contents lists available at ScienceDirect

Energy Economics

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

Distributional modeling and short-term forecasting of electricity prices by Generalized Additive Models for Location, Scale and Shape

Francesco Serinaldi *

School of Civil Engineering and Geosciences, Newcastle University, Newcastle Upon Tyne, NE1 7RU, UK
Willis Research Network, London, UK

ARTICLE INFO

Article history:
Received 21 March 2010
Received in revised form 10 February 2011
Accepted 1 May 2011
Available online 8 May 2011

JEL classification:
C01
C18
C32
C52
C53
Q40

Keywords:
Electricity market price
Point forecasting
Interval forecasting
GAMLSS models
Liberalized energy markets

ABSTRACT

In the context of the liberalized and deregulated electricity markets, price forecasting has become increasingly important for energy company's plans and market strategies. Within the class of the time series models that are used to perform price forecasting, the subclasses of methods based on stochastic time series and causal models commonly provide point forecasts, whereas the corresponding uncertainty is quantified by approximate or simulation-based confidence intervals. Aiming to improve the uncertainty assessment, this study introduces the Generalized Additive Models for Location, Scale and Shape (GAMLSS) to model the dynamically varying distribution of prices. The GAMLSS allow fitting a variety of distributions whose parameters change according to covariates via a number of linear and nonlinear relationships. In this way, price periodicities, trends and abrupt changes characterizing both the position parameter (linked to the expected value of prices), and the scale and shape parameters (related to price volatility, skewness, and kurtosis) can be explicitly incorporated in the model setup. Relying on the past behavior of the prices and exogenous variables, the GAMLSS enable the short-term (one-day ahead) forecast of the entire distribution of prices. The approach was tested on two datasets from the widely studied California Power Exchange (CalPX) market, and the less mature Italian Power Exchange (IPEX). CalPX data allow comparing the GAMLSS forecasting performance with published results obtained by different models. The study points out that the GAMLSS framework can be a flexible alternative to several linear and nonlinear stochastic models.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

In the liberalized electricity markets, the market clearing price is established through one-sided auctions (power pool-type market) or two-sided auctions (power exchange-type market) as the intersection of the supply curve and the estimated demand or demand curve, respectively. Usually, the clearing price is determined the day before delivery by means of 24 or 48 auctions, one for each hour or half an hour of the following day. Therefore, a reliable price forecast plays an important role in the bidding strategies of generator firms as well as of distribution companies, traders and large consumers. According to the planning horizon, price forecasts are classified as long-term (years), medium-term (monthly), and short-term (from few hours to few days), which are particularly useful in the day-to-day operations carried out in the auction-based day-ahead spot market (Bunn, 2000; Weron and Misiorek, 2006). As pointed out by Aggarwal et al. (2009), electricity price time series exhibit patterns more complex than load sequences, and some characteristics (such as non constant mean and variance, multiple strong seasonality, and calendar effect) that can be attributed to the features of the electricity market (non-storability of electricity, inelasticity of the short-term demand, wide spectrum of costs, oligopolistic behavior of the generators) and distinguish electricity from other types of commodities (see also Blumsack et al., 2002; Blumsack et al., 2006; Bosco et al., 2007).

Weron and Misiorek (2006) and Weron (2006) classified the models available in the literature for the electricity price forecasting in six broad classes and focus on statistical methods as the best suited for the short-term forecasting. More recently, Aggarwal et al. (2009) classified price forecasting methods in three wide groups, namely, game theory, simulation, and time series models. The first group aims at mimicking the bidding strategies of the market participants, the second one simulates the physical phenomenon related to the actual dispatch with system requirements and constraints, whereas the last group describes the historical pattern of the prices involving sometimes exogenous variables, such as loads and weather variables. In their review, Aggarwal et al. (2009) considered 47 works, whose results are grouped according to several criteria, namely: (1) type of model, (2) time horizon for prediction, (3) input variables used, (4) output variables, (5) analysis of results, (6) sample size used for the analysis, (7) preprocessing procedures, and (8) model architecture. Focusing on time series models,

* Tel.: +44 0 191 222 5956.
E-mail address: francesco.serinaldi@ncl.ac.uk.

0140-9883/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.eneco.2011.05.001

there is no evidence that one approach clearly outperforms the others. The performance of the models depends on the properties of the data on hand (market maturity, volatility, regulation, time span, etc.), and is usually assessed by comparing point forecasts (i.e., price profiles, which are the most common output of the short-term forecasting models) with the actual clearing prices.

Aggarwal et al. (2009) recognized that only four papers out of 29 dealing with time series models provide confidence intervals or the probability density function associated with price profiles (see Tables 3 and 5 of their paper). Weron (2006) also pointed out the lack of extensive literature on interval forecasting of electricity prices. From 2006, only few authors have complemented the stochastic point forecasts with confidence intervals (e.g., Misiorek and Weron, 2006; Misiorek et al., 2006; Nogales and Conejo, 2006; Weron and Misiorek, 2008). However, the interval forecasting should always be performed to provide a picture of the estimation uncertainty and reliability. Interval forecasting is usually carried out by assuming that prediction errors (i.e., the differences between point forecasts and actual clearing prices) are normally distributed. As an alternative, quantiles with a given probability can be extracted from the empirical distribution of the prediction errors. In other words, the expected price forecast resulting from the point estimation is commonly complemented by confidence intervals deduced from the distribution of the forecasting residuals.

In this study, we assume that the elements of an electricity price time series are realizations of a nonstationary distribution function, the parameters of which change dynamically according to intra-day, weekly, annual seasonality and other possible explanatory variables. Since the position, scale, and shape parameters of a suitable multi-parameter distribution are linked to its expectation, variance, skewness and kurtosis coefficients, a proper choice of the rules driving the change of the parameters' values from a time instant to the following one allows obtaining a price distribution whose expected value represents the time-varying point estimate (forecast), and the time-varying scale and shape parameters can account for hetero-skedastic and leptokurtic (or platykurtic) behavior. Generalized Additive Models for Location, Scale and Shape (GAMLSS) proposed by Rigby and Stasinopoulos (2005) are a well-suited framework to perform this type of analysis because they allow the forecasting of the entire day-ahead distribution and not just the expectation and approximate confidence intervals.

The remainder of the paper is organized as follows. In Section 2, the GAMLSS rationale, inferential aspects and model setup are introduced. Section 3 briefly describes structure and setup of the benchmark models used to compare the GAMLSS performance. Section 4 introduces the criteria used to assess the performence of the forecasting methods. In Section 5, two datasets are analyzed: the California Power Exchange (CalPX) market prices for which results from a number of models are well documented in the literature, and the data from the Italian Power Exchange (IPEX) market, which are of specific interest in the present study. Finally, conclusions close the study.

2. GAMLSS modeling

GAMLSS were proposed by Rigby and Stasinopoulos (2005) as a general modeling framework extending the classical Generalized Additive Models (e.g. Hastie and Tibshirani, 1990), Generalized Linear Models (e.g. Nelder and Wedderburn, 1972; McCullagh and Nelder, 1989), Generalized Linear Mixed Models (e.g. McCulloch, 1997; McCulloch, 2003), and Generalized Additive Mixed Models (e.g. Fahrmeir and Lang, 2001). The underlying rationale of such type of models is that the quantity of concern (the wholesale electricity prices P_t, in the present case) is considered as a response variable whose distribution function varies dynamically according to the values assumed by some explanatory variables (e.g., past price history, electricity loads, and weather variables such as temperature). This dependence is formalized via functional (systematic) relationships between the distribution parameters and the explanatory variables. The advantages of using GAMLSS instead of the other above-mentioned approaches are that (1) the response variable is not restricted to follow a distribution from the exponential family (e.g., Gaussian distribution) allowing for general distribution functions (e.g., highly skewed and/or kurtotic continuous and discrete distributions); and (2) the systematic part allows modeling not only the location parameter (related to the mean), but also scale and shape parameters (related to the dispersion, skewness, and kurtosis) as linear and/or nonlinear, parametric and/or nonparametric additive functions of covariates and/or random effects (e.g., Rigby and Stasinopoulos, 2005; Stasinopoulos and Rigby, 2007).

2.1. Introduction to GAMLSS theory

This section briefly introduces the theory behind GAMLSS, referring the reader to Rigby and Stasinopoulos (2005), Stasinopoulos and Rigby (2007), and references therein for a more comprehensive theoretical presentation. Denoting Y the response variable, for the GAMLSS models it is assumed that independent observations y_i, for $i=1,\ldots, n$, have distribution function $F_Y(y_i;\, \boldsymbol{\theta}^i)$ with $\boldsymbol{\theta}^i=(\theta_1^i,\ldots, \theta_p^i)$ a vector of p distribution parameters accounting for position, scale, and shape. Usually p is less than or equal to four, since 1- , 2- , 3- , and 4-parameter families provide enough flexibility for most applications. Given an n length vector of the response variable $\mathbf{y}^T=(y_1,\ldots, y_n)$, let $g_k(\cdot)$, for $k=1,\ldots, p$, be monotonic link functions relating the distribution parameters to explanatory variables and random effects through an additive model given by:

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \mathbf{X}_k\boldsymbol{\beta}_k + \sum_{j=1}^{J_k} \mathbf{Z}_{jk}\boldsymbol{\gamma}_{jk} \tag{1}$$

where $\boldsymbol{\theta}_k$ and $\boldsymbol{\eta}_k$ are vectors of length n, e.g. $\boldsymbol{\theta}_k^T=\{\theta_k^1,\ldots, \theta_k^n\}$, $\boldsymbol{\beta}_k^T=\{\beta_{1k},\ldots, \beta_{J_k}k\}$ is a parameter vector of length J_k, $\mathbf{X}_k$ is a known design matrix of order $n\times J_k$, $\mathbf{Z}_{jk}$ is a fixed known $n\times q_{jk}$ design matrix and $\boldsymbol{\gamma}_{jk}$ is a q_{jk}-dimensional random variable.

In Eq. (1), the linear predictors $\boldsymbol{\eta}_k$, for $k=1,\ldots, p$, are comprised of a parametric component $\mathbf{X}_k\boldsymbol{\beta}_k$ (linear functions of explanatory variables), and additive components $\mathbf{Z}_{jk}\boldsymbol{\gamma}_{jk}$ (linear functions of stochastic variables, also denoted as random effects). GAMLSS involve several important sub-models. In particular, if $\mathbf{Z}_{jk}=\mathbf{I}_n$, where $\mathbf{I}_n$ is an $n\times n$ identity matrix, and $\boldsymbol{\gamma}_{jk}=\mathbf{h}_{jk}=h_{jk}(\mathbf{x}_{jk})$ for all combinations of j and k in model (1), we have the semi-parametric additive formulation of GAMLSS:

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \mathbf{X}_k\boldsymbol{\beta}_k + \sum_{j=1}^{J_k} h_{jk}\left(x_{jk}\right) \tag{2}$$

where h_{jk} is an unknown function (local linear smoothers, splines, etc.) of the explanatory variable $\mathbf{X}_{jk}$ and $\mathbf{h}_{jk}=h_{jk}(\mathbf{x}_{jk})$ is the vector that evaluates the function h_{jk} at $\mathbf{x}_{jk}$. Additive terms in formulation (2) represent smoothing terms that allow for additional flexibility in modeling the dependence of the distribution parameters on the explanatory variables. The reader is referred to Villarini et al. (2009b) and Villarini et al. (2009a) for example applications using linear functions $\mathbf{X}_k\boldsymbol{\beta}_k$ and cubic spline smoothing relationships $h(\cdot)$.

Model fitting and selection are discussed in Stasinopoulos and Rigby (2007). The inference procedure implies the selection of a suitable distribution family of Y, the explanatory variables, the link functions, and the structure of the systematic part (i.e., linear and/or nonlinear, parametric and/or nonparametric additive functions between parameters and covariates). Since the estimation method suggested by Stasinopoulos and Rigby (2007) is based on the maximum likelihood principle, the model selection can be carried out by checking the significance of the fitting improvement in terms of information criteria such as the Akaike Information Criterion (AIC; Akaike, 1974), the Schwarz Bayesian Criterion (SBC; Schwarz, 1978), or the generalized AIC (GAIC; Stasinopoulos and

ISI Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات

ISI Articles