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.

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 $i=1,\dots, n$, have distribution function $F_Y(y_i;\, \boldsymbol{\theta}^i)$ with $\boldsymbol{\theta}^i=(\theta_1^i,\dots, \theta_p^i)$ 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 $\mathbf{y}^T=(y_1,\dots, y_n)$, let $g_k(\cdot)$, for $k=1,\dots, p$, 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. $\boldsymbol{\theta}_k^T=\{\theta_k^1,\dots, \theta_k^n\}$, $\boldsymbol{\beta}_k^T=\{\beta_{1k},\dots, \beta_{J_k}k\}$ 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 $k=1,\dots, p$, 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

Rigby, 2007). Forward, backward, and step-wise procedures can be applied to select the meaningful explanatory variables. Moreover, a number of diagnostic plots to check the fitting performance are discussed in the study by Stasinopoulos and Rigby (2007), which the reader is referred to for more details.

2.2. GAMLSS setup

In this study, we have applied the likelihood-based strategies only for a preliminary selection because we are interested in obtaining a model that minimizes the forecast error rather than the fitting error. This requires a somewhat time-expensive but unavoidable try-and-error approach.

In more detail, we have considered 17 continuous distribution families with two, three, and four parameters among those listed by Stasinopoulos and Rigby (2007). To restrict the possible combinations of distribution families and structures of the systematic part (Eq. 1), the functional relationship between the position parameter of the 17 distributions and the covariates used in the benchmark models (described later) was assumed linear, whereas the scale and shape parameters were assumed as constant values. Therefore, the parameters of these 17 models were estimated by minimizing the maximum likelihood and the models were ranked according to the values of their AIC and SBC. This procedure led to select the 4-parameter Johnson S_u distribution (JSU; Johnson, 1949) with density function:

$$f_Y(y;\mu,\sigma,\nu,\tau) = \frac{\tau}{c\sigma}\frac{1}{(r^2+1)^{1/2}}\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}z^2\right], \tag{3}$$

for $y\in(-\infty,\infty)$, where $\mu\in(-\infty,\infty)$ is a position parameter, $\sigma>0$ is a scale parameter, $\nu\in(-\infty,\infty)$ and $\tau>0$ are shape parameters, and where:

$$z = -\nu + \tau\log\left[r + \left(r^2+1\right)^{1/2}\right], \tag{4}$$

$$r = \frac{y-\left(\mu + c\sigma w^{1/2}\sinh\Omega\right)}{c\sigma}, \tag{5}$$

$$c = \left\{\frac{1}{2}(w-1)[w\cosh(2\Omega)+1]\right\}^{-1/2}, \tag{6}$$

$w=\exp(1/\tau^2)$, and $\Omega=-\nu/\tau$. Z is a standard normal random variable, and it is $E[Y]=\mu$ and $Var[Y]=\sigma^2$. This distribution is appropriate for leptokurtic data. The parameter ν controls the skewness of the distribution with $\nu>0$ denoting positive skewness and $\nu<0$ negative. The parameter τ determines the kurtosis of the distribution, which approaches the Gaussian density function as $\tau\rightarrow\infty$.

After selecting the distribution family, the structure of the model (namely, the functional relationships in Eq. 1) was refined through an iterative procedure by trying several combinations of explanatory variables and functional relationships (e.g., polynomial with different degrees) for all parameters, and assuming the model that yielded the most accurate forecasts according to the performance measures described in Section 4. As mentioned above, the automatic procedures of selection based on likelihood criteria were not applied in this context because we are interested in the "best forecasting" model, which do not always correspond to the "best fitting" one (e.g., over-parameterized models can

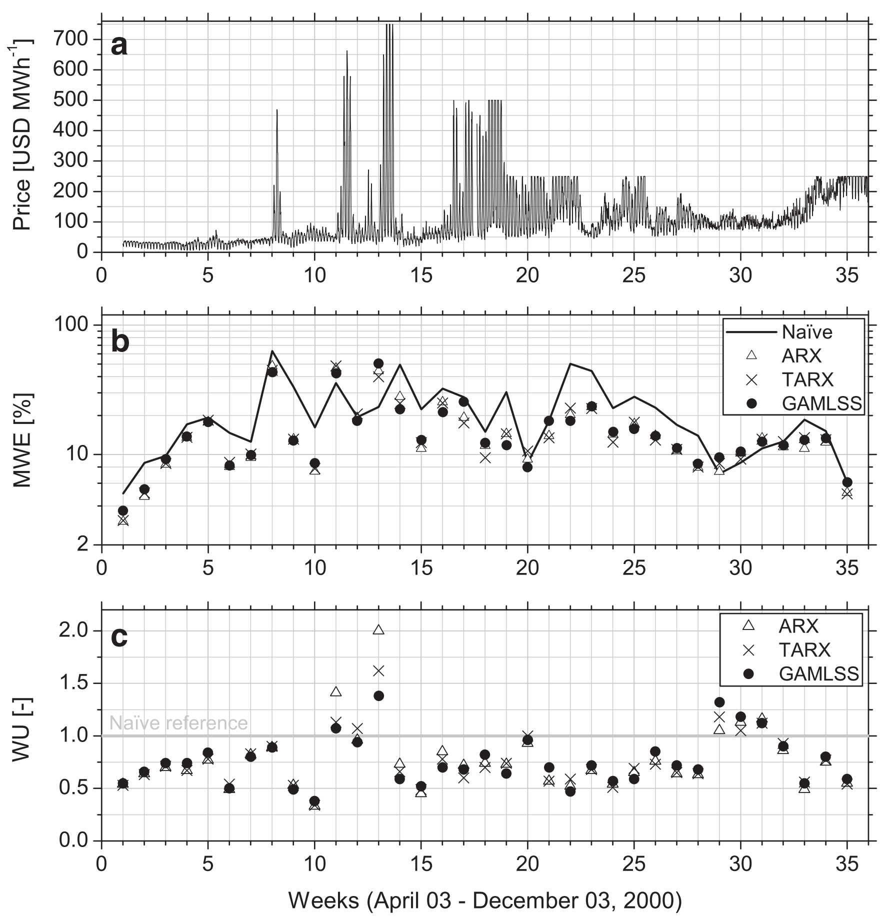

Fig. 1. (a) California Power Exchange (CalPX) market clearing prices for 35 weeks used to test the forecasting methods. (b-c) Mean weekly error (MWE) and weekly Theil's U-statistics (WU) patterns for the tested models (see text for details).

show very small fitting errors but poor forecasting performances). This procedure returned the following model structure.

Since the μ parameter represents the expected price, it was assumed to be a suitable estimate of the point forecast. Hence, the structure of the relationship between μ and the explanatory variables was chosen starting from those used by Misiorek et al. (2006) for the benchmark models describe in Section 3. Namely, we assumed an identity link function $g_1(\cdot)$, and the expected price $E[P_t]$ depending on the prices at the same hour of 1 day before the time t (P_{t-24}), 2 days before t (P_{t-48}), 1 week before t (P_{t-168}), on the minimum of the hourly prices in the previous day (M_t), and on the load forecast or actual purchased load (L_t) according to data availability. Moreover, we used only linear components $\mathbf{X}_k\boldsymbol{\beta}_k$ for the sake of parsimony. Therefore, Eq. 1 for the μ parameter (which corresponds to the theoretical and unobserved JSU price expectation) becomes

$$\begin{aligned} g_1(\theta_1) = \mu = E[P_t] = & \beta_{\mu P,1}P_{t-24} + \beta_{\mu P,2}P_{t-48} + \beta_{\mu P,3}P_{t-168} \\ & + \beta_{\mu M,1}M_t + \sum_{k=1}^{3}\beta_{\mu L,k}(L_t)^k \\ & + \sum_{k=1}^{7} d_{d,k}D_{d,k} + \sum_{k=1}^{24} d_{h,k}D_{h,k}, \end{aligned} \tag{7}$$

where $d_{d,k}$, $k=1$, ..., 7, and $d_{h,k}$, $k=1$, ..., 24, are coefficients of the dummy variables $D_{d,\,k}$ and $D_{h,\,k}$, which account for weekly and daily periodicity by distinguishing the price patterns of each day and each hour. The μ parameter is linked to loads through a 3-order polynomial.

The σ parameter (which corresponds to the theoretical and unobserved JSU price standard deviation) was assumed to be dependent on the previous-day price P_{t-24}, and on the absolute value of rate of price change $R_{t-24}=|\log P_{t-24}-\log P_{t-48}|$, which summarizes the daily price volatility. Moreover, a logarithmic link function $g_2(\cdot)$ is used to guarantee the fulfillment of the condition $\sigma>0$. Dummy variables are introduced as well, resulting in the relationship:

$$\begin{aligned} g_2(\theta_2) = \log\sigma = \log\sqrt{Var[P_t]} = & \beta_{\sigma P,1}P_{t-24} + \beta_{\sigma R,1}R_{t-24} \\ & + \sum_{k=1}^{7} d_{d,k}D_{d,k} + \sum_{k=1}^{24} d_{h,k}D_{h,k}. \end{aligned} \tag{8}$$

Finally, the shape parameters were set as constant for the sake of simplicity and parsimony. Therefore, ν and τ are:

$$g_3(\theta_3) = \nu = \beta_{\nu,0}, \tag{9}$$

$$g_4(\theta_4) = \log\tau = \beta_{\tau,0}. \tag{10}$$

In this study, all the GAMLSS calculations were performed in R (R Development Core Team, 2009) using the freely available gamlss package (Stasinopoulos et al., 2007).

Table 1
Mean weekly error (MWE) for each week of the testing period of CalPX data. The relative calm weeks 1–10 are distinguished from the more volatile weeks 11–35. The number of weeks in which a model is selected as the most accurate ("best") along with the mean deviation from the best (m.d.f.b.) model is also shown. Bold character denotes the best result for each row of the table.

Week	Naïve	ARX	TARX	GAMLSS
1	5.00	**3.03**	3.09	3.68
2	8.62	**4.71**	5.04	5.40
3	9.74	**8.37**	8.52	9.20
4	17.14	**13.51**	13.56	13.82
5	19.31	**17.82**	18.45	17.89
6	14.7	**8.04**	8.69	8.19
7	12.56	**9.43**	10.07	9.87
8	62.97	48.15	44.77	**43.30**
9	33.22	13.11	13.12	**12.77**
10	16.23	**7.39**	7.77	8.57
11	**35.59**	46.23	48.34	42.21
12	19.41	19.23	20.63	**18.20**
13	**23.31**	44.17	39.81	50.51
14	49.47	27.99	24.8	**22.33**
15	22.37	**11.11**	12.35	12.90
16	32.35	25.41	24.96	**21.23**
17	27.74	19.26	**17.53**	25.56
18	15.00	11.71	**9.46**	12.29
19	30.42	14.47	14.45	**11.86**
20	8.60	9.18	10.66	**7.97**
21	18.22	13.91	**13.44**	18.11
22	50.33	20.28	22.88	**18.15**
23	44.17	23.27	**22.65**	23.49
24	22.86	14.30	**12.47**	14.94
25	27.90	17.28	17.74	**15.66**
26	22.99	13.97	**12.91**	13.87
27	16.98	**10.65**	10.96	11.24
28	13.96	**7.93**	8.01	8.46
29	**7.11**	7.36	8.52	9.51
30	**8.66**	10.22	9.21	10.54
31	**11.12**	13.35	13.13	12.56
32	12.62	**11.43**	12.62	11.83
33	18.57	**11.09**	13.53	12.96
34	15.15	**12.40**	13.04	13.37
35	6.09	5.07	**4.96**	6.09
best 1–10	0	**8**	0	2
m.d.f.b. 1-10	7.11	0.52	0.47	**0.43**
best 11–35	5	6	**7**	**7**
m.d.f.b. 11–35	8.07	2.48	**2.39**	2.66

3. Benchmark statistical models

The performance of the GAMLSS was tested against some statistical models used as reference. In particular, classical linear AutoRegressive (AR) models with exogenous variables (ARX), ARX with residuals modeled by Generalized AutoRegressive Conditional Heteroskedastic (GARCH; Bollerslev, 1986) model, and the nonlinear Threshold Auto-Regressive (TAR) model with exogenous variables (TARX) were applied, as these approaches are linear and nonlinear, and merge simplicity, flexibility and good performances in price forecasting.

Of course, other models could be considered for the sake of comparison, such as Markov regime switching models, jump-diffusion models and methods based on artificial intelligence (see e.g., Aggarwal et al., 2009; Serati et al., 2008; Weron, 2006). However, we excluded the artificial intelligence-based models because they fall in a different (data-driven) subclass of time series models, whereas the Markov regime switching models were found more accurate that other linear models for spiky periods but they are generally outperformed in the out-of-sample forecasting (e.g., Bessec and Bouabdallah, 2005; Dacco and Satchell, 1999; Haldrup and Nielsen, 2006; Kosater and Mosler, 2006; Misiorek et al., 2006; Weron, 2006). Jump-diffusion models were not applied as they are mainly designed to mimic the statistical properties of electricity prices (commonly at the daily time scale) rather than providing accurate forecasts of hourly prices. Therefore, the simplicity and analytical tractability of this type of models are an advantage in the context of derivative evaluation, but may be too restrictive for short-term price forecasting (e.g., Weron, 2006, pp 104–105).

Since stochastic models (AR, TAR and their extensions) have been extensively applied and compared by Misiorek et al. (2006) and Weron (2006) (see also references therein for further studies and examples), the basic notation is recalled in the Appendix, whereas the models' setup for the data on hand is described in this section.

3.1. Setup of AR and AR-GARCH models

AR models and their extensions were applied to CalPX hourly data (described in Section 5.1.1) and compared by Misiorek et al. (2006). Since these results have been used as reference to test the GAMLSS performance, the same setup adopted by Misiorek et al. (2006) was used for both case studies presented in Section 5. The structure of the AR model was assumed to be as follows:

$$P_t = \phi_1 P_{t-24} + \phi_2 P_{t-48} + \phi_3 P_{t-168} + \phi_4 M_t + d_1 D_{Mon} + d_2 D_{Sat} + d_3 D_{Sun} + \varepsilon_t, \quad (11)$$

where P_{t-24}, P_{t-48}, P_{t-168}, and M_t have been specified in the previous section; d_1, d_2, and d_3 are coefficients of the dummy variables D_{Mon}, D_{Sat}, and D_{Sun}, which account for the weekly seasonality by distinguishing the price patterns of Monday, Saturday, and Sunday from those of the other days. The corresponding ARX model is obtained by adding an exogenous variable, namely the load forecasts, through a further term $\psi_1 L_t$, where ψ_1 is the coefficient of the load forecast L_t. The AR-GARCH and ARX-GARCH models are obtained by assuming that residuals are not white noise, but are given by:

$$\xi_t = \varepsilon_t \sigma_t \quad \text{with} \quad \sigma_t^2 = \alpha_0 + \alpha_1 \xi_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \quad (12)$$

These structures were selected by Misiorek et al. (2006) for CalPX hourly data via the trial-and-error approach, looking for the "best" structure in terms of forecasting performance in the first week of the testing period. Both AR(X) and AR(X)-GARCH models were calibrated using the prediction error estimate, which search the parameters' set that minimizes the differences between observations and model output (for Gaussian residuals, this method coincides with the maximum likelihood approach). Moreover, the benchmark models are fitted to 24 hourly series assuming a variable segmentation strategy (e.g., Weron and Misiorek (2006)).

3.2. Setup of TAR models

The structure of the two components of TAR and TARX models in Eq. (A.4) was assumed equal to that of the AR and ARX, and the parameters were estimated by the prediction error method, applying a variable segmentation strategy. The identification of the threshold variable and level was performed by trial-and-error. Namely, Misiorek et al. (2006) tried different choices of the switching variable u_t and the corresponding threshold T, concluding that the difference between the mean prices for the day before the date t ($\overline{P}_{t-1d}$) and eight days before t ($\overline{P}_{t-8d}$) with the arbitrary threshold $T = 0$ yields better results than other choices for the CalPX prices in the out-of-sample testing period. Weron (2006) denoted TAR and TARX as TAR-P and TARX-P to distinguish them from analogous models TAR-L and TARX-L, for which the threshold variable is selected as the difference of the mean loads (instead of prices) for one and eight days before t. It is worth noting that these models imply an implicit exogenous variable because the loads appear in the switching

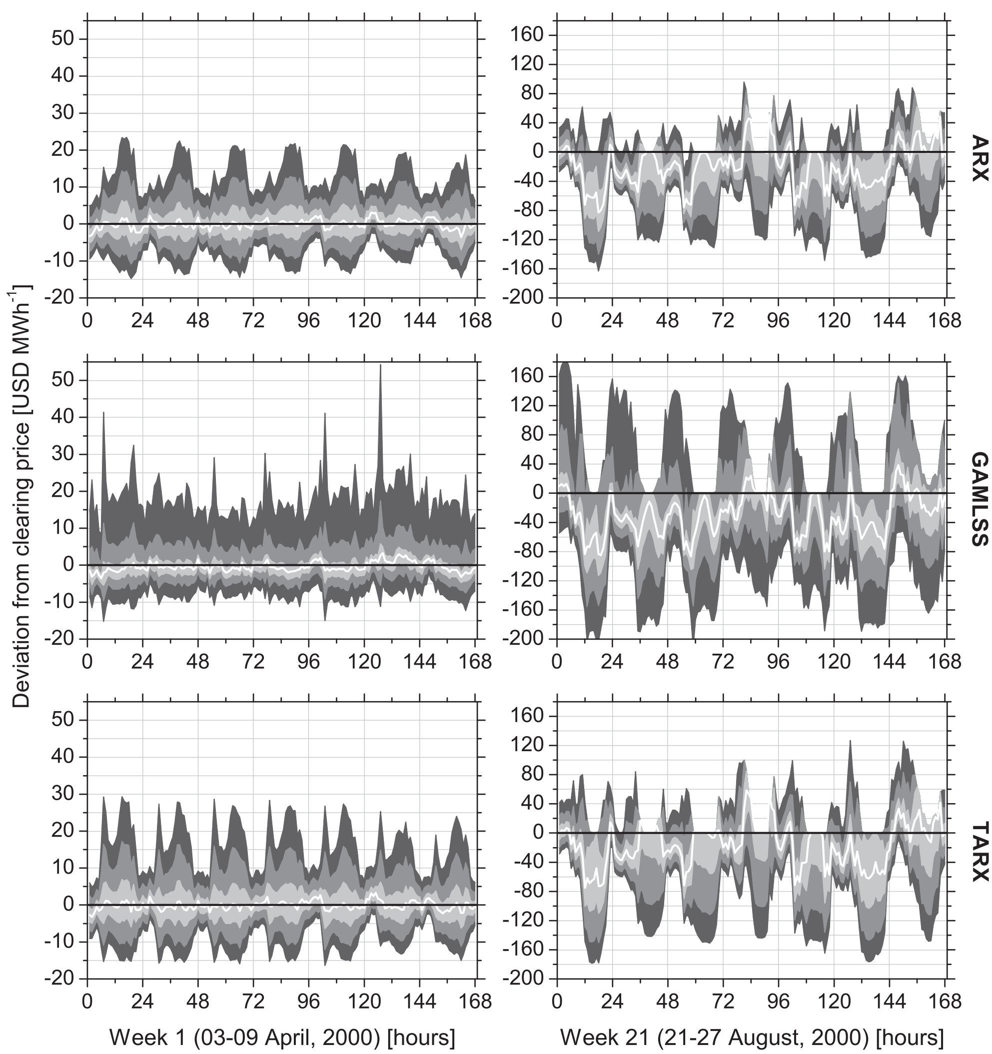

Fig. 2. Deviations of the point forecasts and limits of the 50%, 90%, and 99% two-sided CIs from the market clearing price for the 1st and 21st week of the testing period of CalPX data.

rule. Since forecast loads are not always available, TAR-L and TARX-L were not applied; however, results for CalPX data can be found in the work by Weron (2006, pp. 133–136). Therefore, only TAR-P and TARX-P with $u_t = (\overline{P}_{t-1d} - \overline{P}_{t-8d})$ and $T=0$ were used in this study. For further details about the AR-based benchmark models, the readers are referred to Misiorek et al. (2006) and Weron (2006, pp. 109–136). All calculations referring to reference models were performed by the MFE toolbox version 1.0.1 for Matlab^{R} made available by Dr. Weron at the web site http://www.ioz.pwr.wroc.pl/pracownicy/weron/MFE.htm.

3.3. Naïve forecast

The point forecasts were also compared with those resulting from the naïve or similar-day approach based on the following rules: Monday prices are similar to the Monday prices of the previous week, and the same applies to Saturday and Sunday; instead, Tuesday is assumed to be similar to the previous Monday, and the same applies to Wednesday, Thursday, and Friday.

4. Assessment of model performance

Several measures and indices are available to assess the forecast accuracy (e.g., Hyndman and Koehler, 2006; Makridakis et al., 1998). Weron (2006, pp. 107–109) lists some measures based on linear and quadratic norms to assess daily and weekly forecast errors. Even though the different indices can yield slightly different results, we focused on Mean Week Error (MWE) defined as:

$$\mathrm{MWE} = \frac{1}{168}\sum_{h=1}^{168}\frac{\left|P_h - \hat{P}_h\right|}{\overline{P}_{168}}, \tag{13}$$

where P_h is the market clearing price, $\hat{P}_h$ denotes the price forecast, and $\overline{P}_{168}$ is the weekly mean price. Since it is not unusual that naïve methods

Table 2
Percentage of data falling outside the CIs with nominal coverage probability equal to 50%, 90%, and 99% for the whole testing periods of CalPX and IPEX data. For CalPX data, the relative calm weeks 1–10 are distinguished from the more volatile weeks 11–35. Bold character denotes the models' exceedance probabilities closest to the nominal values for each dataset and class of probability.

Data	Week	Model	50%	10%	1%
CalPX	1–10	ARX	41.67	13.57	5.60
		AR-GARCH	41.85	12.08	5.12
		TARX	37.32	**11.31**	4.58
		GAMLSS	**55.65**	14.70	**3.04**
-	11–35	ARX	45.95	13.45	5.48
		AR-GARCH	42.86	13.45	5.60
		TARX	38.19	**8.86**	3.17
		GAMLSS	**50.11**	8.43	**0.57**
IPEX	1–52	ARX	**53.24**	15.68	4.90
		ARX-GARCH	65.16	25.98	8.87
		TARX	40.37	**9.21**	2.59
		GAMLSS	54.15	13.15	**1.98**

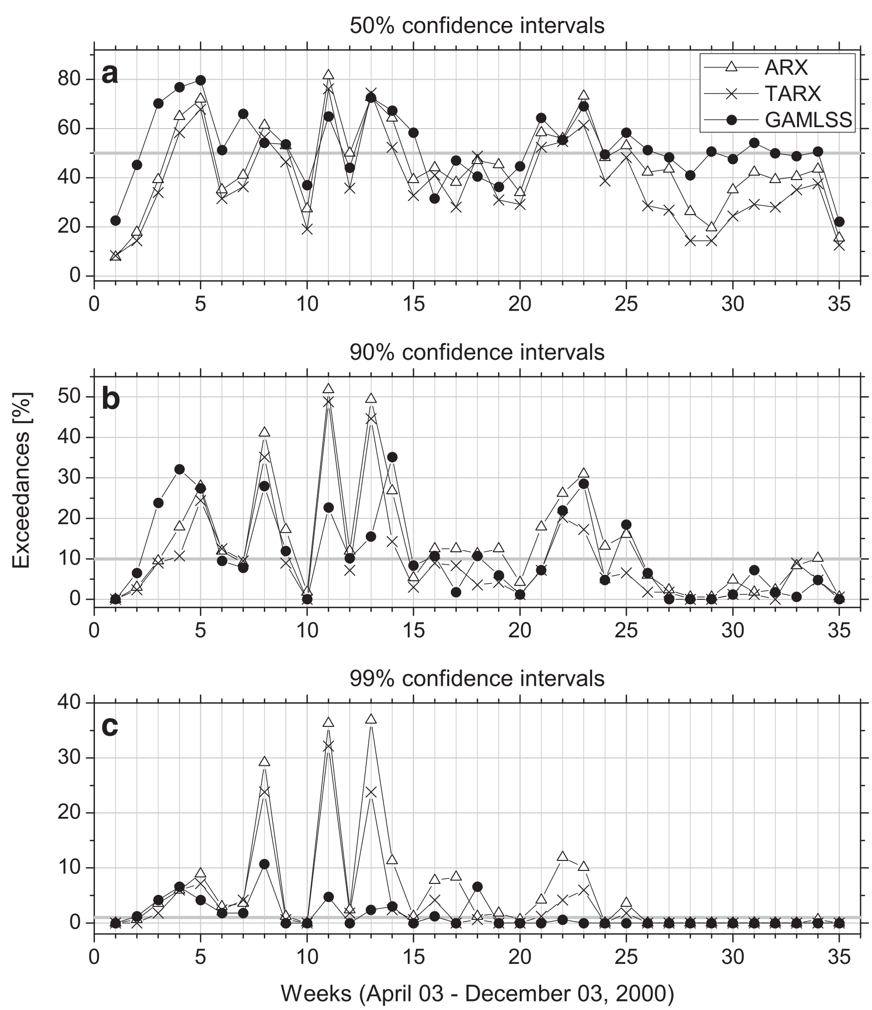

Fig. 3. Percentage of data falling outside the CIs with nominal coverage probability equal to (a) 50%, (b) 90%, and (c) 99% for each week of the testing period of CalPX data.

perform better than more complex not well-calibrated models, we also used the Theil's U-statistic, which provides a direct comparison of complex methods with naïve approaches via squared norms giving more weight to large errors than to small errors (Makridakis et al., 1998, pp. 48–50). The weekly U-statistic (WU) is defined as follows:

$$\mathrm{WU} = \sqrt{\frac{\sum_{h=1}^{168}\left(\frac{\hat{P}_h - P_h}{\ddot{P}_h}\right)^2}{\sum_{h=1}^{168}\left(\frac{P_h - \ddot{P}_h}{\ddot{P}_h}\right)^2}}, \tag{14}$$

where $\ddot{P}_h$ is the naïve forecast. The values of WU allow an easy comparison of formal and naïve approaches:

-if WU = 1: the tested forecasting method performs as well as the naïve forecast.
-if WU < 1: the tested forecasting method performs better than the naïve one. The smaller the value of WU, the better is the model compared to the naïve.
-if WU > 1: the tested forecasting method performs worse than the naïve forecast.

5. Data analyses

5.1. Analysis of California Power Exchange (CalPX) electricity prices

5.1.1. CalPX data and preprocessing

The GAMLSS and the benchmark models mentioned in Sections 2 and 3 were first applied to CalPX data since Misiorek and Weron (2006) and Weron (2006) provided results of one-day ahead point and interval forecasting for all benchmark models and for an extensive out-of-sample testing period of 35 weeks. Hourly market clearing prices and load forecasts can be retrieved from the web sites of University of California Energy Institute (UCEI; http://www.ucei.berkeley.edu) and CAlifornia Independent System Operator (CAISO) (http://oasis.caiso.com); however, we used the data preprocessed for missing and "doubled" data corresponding to changes to and from summer time, provided by Dr. Weron (http://www.ioz.pwr.wroc.pl/pracownicy/weron/MFE.htm) and described in detail by Weron (2006, pp. 114–116). All models (GAMLSS and benchmark) were applied to prices and loads preprocessed by logarithmic transformation to obtain a more stable variance. Models were calibrated for the period July 5, 1999–April 2, 2000, whereas the period April 3–December 3, 2000 was used for out-of-sample testing (Fig. 1a). The forecasting of the 24 hourly prices of a given day was performed by estimating the model parameters by all data up to the day of interest. This approach (involving an increasing sample for parameter estimation) was found to perform better than a sliding windows method in terms of forecasting accuracy for the

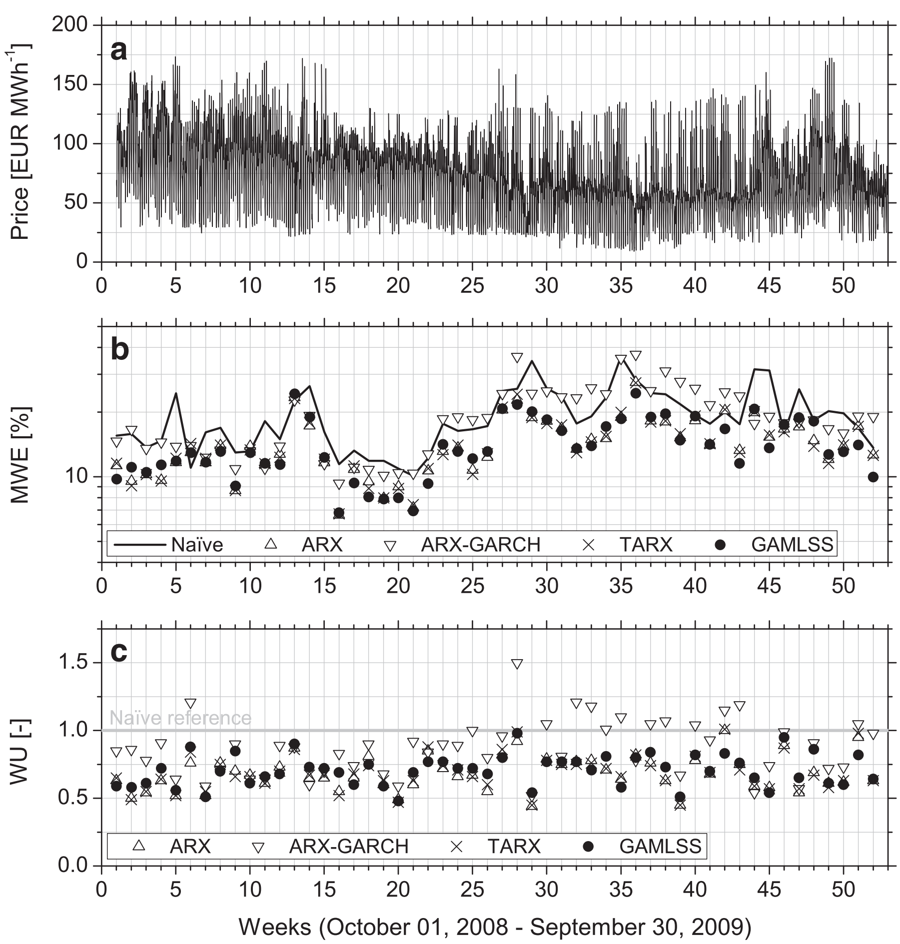

Fig. 4. (a) Italian Power Exchange (IPEX) market clearing prices for the 52 weeks used to test the forecasting methods. (b-c) MWE and WU performance statistics for the tested models (see text for details).

benchmark models (Misiorek and Weron, 2006). However, since GAMLSS were found not to be sensitive to this choice, a sliding windows approach was used for these models.

5.1.2. CalPX results

Results of the point forecasting are summarized in Fig. 1b-c and Table 1. Like Misiorek et al. (2006), we distinguished the rather calm first 10 weeks and the more volatile weeks 11–35. Results for AR(X)-GARCH models are not reported because these methods showed a worse performance relative to AR(X) and TAR(X). Fig. 1b shows that the pattern of the GAMLSS MWE is close to those of ARX and TARX models. The naïve method outperforms ARX and TARX six times out of 35 in terms of WU index, while GAMLSS five times. The values of WU are almost always less than 0.8 denoting that the tested models can give a significant improvement in the short-term forecasting. A closer look at the results is provided in Table 1. For the calm first 10 weeks, the ARX model performs better than the others in terms of MWE; however, the GAMLSS approach shows the minimum mean deviance from the best model for each week, meaning that it performs close to the best model even for cases in which it is not the best. For the volatile weeks 11–35, TARX and GAMLSS slightly outperform the naïve and ARX methods, but an overall winner does not seem evident. ARX, TARX, and GAMLSS exhibit similar values of the mean deviance from the best model for a specific week, whereas the naïve method strongly differs from the best model when it is not the best.

As previously mentioned, the GAMLSS allow modeling the full dynamic distribution of prices and therefore the time-varying quantiles, which can provide interval forecasts. On the other hand, confidence intervals (CIs) for AR(X), AR(X)-GARCH, and TAR(X) models are computed by using the empirical distribution of the one-step ahead forecast errors or by its Gaussian approximation (Weron, 2006, p. 136). The accuracy of the CIs is checked by comparing the nominal coverage probability (e.g., 50%, 90%, and 99%) with the actual one. Fig. 2 shows the deviations of the point forecasts and limits of the 50%, 90%, and 99% two-sided CIs from the market clearing price for the 1st and 21st week of the testing period. Panels referring to ARX and TARX models reproduce those shown by Misiorek et al. (2006). Focusing on the 1st (rather calm) week, GAMLSS 50% CIs correctly include the market clearing price a percentage of times, whereas ARX and TARX 50% CIs incorrectly enclose all prices. The GAMLSS 99% CIs are wider than those of ARX and TARX according to the leptokurtic nature of JSU distribution. Moreover, since the scale parameter (the variance) of the JSU distribution was assumed to be dependent on the previous-day return as well as previous-day price, GAMLSS CIs do not show the daily cyclicity of ARX and TAR CIs. For the more volatile 21st week (panels on the right of Fig. 2), the intervals are wider than CIs of the 1st week owing to the greater uncertainty. For some hours, the point forecasts and the upper limit of CIs collapse to zero because of the price cap to 250 USD MWh^{-1} (see Fig. 1a) introduced by CAISO during the California market crisis occurred in 2000. The upper limit of the GAMLSS CIs is almost always greater than that of the corresponding ARX and TARX. In other words, the dynamic JSU distribution accounts for the asymmetry of the price distribution and recognizes that the forecasting uncertainty mainly affects possible high prices, whereas lower limits can be better predicted. Fig. 3 summarizes the actual probabilities that prices fall outside the 50%, 90%, 99% CIs for the 35 weeks and all considered models. Any model seems to be able to closely reproduce the nominal coverage probability. However, focusing on the 50% CIs, GAMLSS perform rather well for weeks 24 to 34, and better than the other models for 99% CIs. Results in Table 2 point out that the actual coverage probability of GAMLSS CIs is on an average the closest to the nominal one for both calm and volatile periods (weeks 1–10, 11–35), and for 50% and 99% CIs. For 90% CIs, TARX outperforms ARX and GAMLSS; however, the GAMLSS coverage probability is very close to that of TARX for the volatile weeks, whereas it shows the worst performance in the calm weeks.

5.2. Analysis of Italian Power Exchange (IPEX) electricity prices

5.2.1. IPEX data and preprocessing

The analyses carried out on CalPX data were repeated for hourly data from the Italian electricity market. This market has operated since April 1, 2004 as a pool market. Demand-side bidding was introduced for a small number of eligible buyers on January 1, 2005, and was extended to all buyers on July 1, 2007 to account for European Directive 2003/54/EC (Bosco et al., 2007). Data are available at the web site of the independent market operator named Gestore Mercati Energetici (GME; www.mercatoelettrico.org). For data spanning from April 1, 2004 to January 15, 2006, Bosco et al. (2007) found that a compound model involving a time regressive component, a

Table 3
Mean weekly error (MWE) for each week of the testing period of IPEX data. The number of weeks in which a model is selected as the most accurate ("best") along with the mean deviation from the best (m.d.f.b.) model is also shown. Bold character denotes the best result for each row of the table.

Week	Naïve	ARX	ARX-GARCH	TARX	GAMLSS
1	15.52	11.27	14.64	11.58	**9.76**
2	15.80	9.53	16.62	**9.08**	11.07
3	13.68	**10.23**	13.53	**10.23**	10.50
4	14.70	9.62	14.49	**9.55**	11.35
5	24.43	**11.65**	13.86	12.21	11.83
6	**11.01**	12.97	13.95	14.30	12.96
7	16.11	**11.62**	12.31	12.19	11.69
8	16.92	13.98	13.32	14.03	**13.14**
9	12.96	**8.62**	10.90	8.63	9.08
10	13.12	13.85	13.20	13.34	**12.95**
11	18.17	11.39	**10.96**	11.51	11.56
12	14.95	12.73	13.89	12.33	**11.43**
13	**22.66**	23.35	22.78	23.09	24.30
14	26.39	**17.25**	19.27	17.83	18.94
15	16.16	11.63	**11.51**	12.09	12.34
16	11.45	**6.64**	9.35	6.73	6.81
17	13.23	10.83	11.07	11.25	**9.36**
18	11.85	9.49	10.82	8.82	**8.06**
19	11.87	7.94	10.15	7.98	**7.89**
20	10.92	8.98	10.44	8.82	**7.98**
21	10.11	7.17	10.39	7.43	**6.93**
22	12.23	10.67	12.76	10.94	**9.28**
23	17.67	13.15	18.63	**12.68**	14.16
24	16.35	13.60	19.01	14.05	**13.14**
25	16.63	10.74	18.45	**10.27**	12.14
26	17.22	**12.33**	18.85	12.80	13.11
27	25.07	**20.46**	24.38	21.07	20.75
28	25.64	22.73	36.26	24.19	**21.70**
29	34.57	**18.85**	24.43	18.88	20.04
30	25.70	18.17	25.10	**17.70**	18.42
31	23.70	16.80	23.52	17.42	**16.38**
32	17.69	13.44	23.32	**12.94**	13.53
33	19.11	14.90	25.90	14.53	**13.92**
34	23.35	**15.08**	24.32	15.44	17.18
35	36.34	18.90	35.57	19.95	**18.58**
36	28.22	27.47	37.12	27.34	**24.46**
37	24.51	18.19	25.13	**17.98**	19.02
38	24.17	**17.99**	31.06	18.21	19.66
39	21.72	15.20	27.74	15.85	**14.78**
40	19.37	**18.24**	25.76	19.19	19.14
41	17.67	**14.02**	21.62	14.25	14.15
42	20.05	20.32	24.84	20.58	**16.74**
43	17.58	13.27	23.77	12.99	**11.55**
44	31.63	19.90	**17.69**	20.10	20.69
45	31.24	15.32	19.11	15.67	**13.61**
46	**15.93**	16.70	17.82	16.17	17.50
47	25.56	**17.06**	17.95	17.73	18.82
48	18.33	14.74	18.38	**13.85**	18.14
49	20.18	12.06	16.75	**11.54**	12.70
50	19.75	13.81	16.00	14.07	**13.06**
51	16.85	17.19	19.16	17.18	**14.03**
52	13.66	12.65	19.04	12.58	**9.97**
best	3	14	3	10	**23**
m.d.f.b.	5.68	0.78	5.44	0.90	**0.70**

periodic AR(5) term, and GARCH residuals performs better than other models in terms of AIC and SBC. In this study, we focus on the forecasting ability of the models used in the previous section using the testing period October 1, 2008–September 30, 2009 (Fig. 4a). Unlike CAISO, GME does not provide load forecasts, which therefore are replaced with the previous-day actual purchased loads. Similar to CalPX data, "doubled" values corresponding to changes to and from summer time were replaced by the arithmetic average of the two values for the "doubled" hour, and all models (GAMLSS and benchmark) were applied to log-transformed prices and loads.

5.2.2. IPEX results

The IPEX price for the testing period does not show evident changes in the volatility. The pattern of the MWE (Fig. 4b) points out that the ARX-GARCH returns higher errors than the other models. Fig. 4c highlights that ARX-GARCH is outperformed in several cases by the naïve approach. In more detail, GAMLSS yield the best results in terms of MWE for 23 weeks out of 52 (see Table 3). GAMLSS are followed by the simple ARX and TARX. The score of TARX can be ascribed to the reasonable lack of more than one regime in the IPEX prices. The mean deviation from the best model for each week confirms that GAMLSS yield quite accurate forecasts also when they are not the best method. Fig. 5 shows the deviation of point estimate and 50%, 90%, and 99% two-sided CIs from the market clearing price for two example weeks: the 13th (December 24–30, 2008), and 16th (January 14–20, 2009). The 13th week represents the first week of the Christmas holidays (December 24–January 6) characterized by a typical sudden decrease of the electricity consumption (Bosco et al., 2007). Therefore, the models exhibit high MWE because this information is not included in the model setup. On the other hand, the 16th is a business week, in which the tested models yield small MWE, denoting an accurate point estimate. In this case, GAMLSS CIs are closer to AR CIs than to TARX CIs. As for CalPX data, the width of CIs reflects the leptokurtic behavior of JSU distribution compared to ARX error distribution. Focusing on the 16th week, the TARX 50% CIs include the clearing price almost always instead of the expected 50% of times. Fig. 6 points out that all models reproduce far from accurately the nominal coverage probabilities for the testing period. On an average, GAMLSS give the best result for the 99% CIs (see Table 2), TARX for the 90% CIs, and ARX for the 50% CIs. However, in the last two cases, GAMLSS return the second best result after ARX and TARX. In particular, for the 50% CIs, the exceedance probability of GAMLSS (54.15%) is very close to that given by ARX (53.24%).

6. Conclusions

In this study, GAMLSS models proposed by Rigby and Stasinopoulos (2005) have been introduced as an alternative stochastic method to

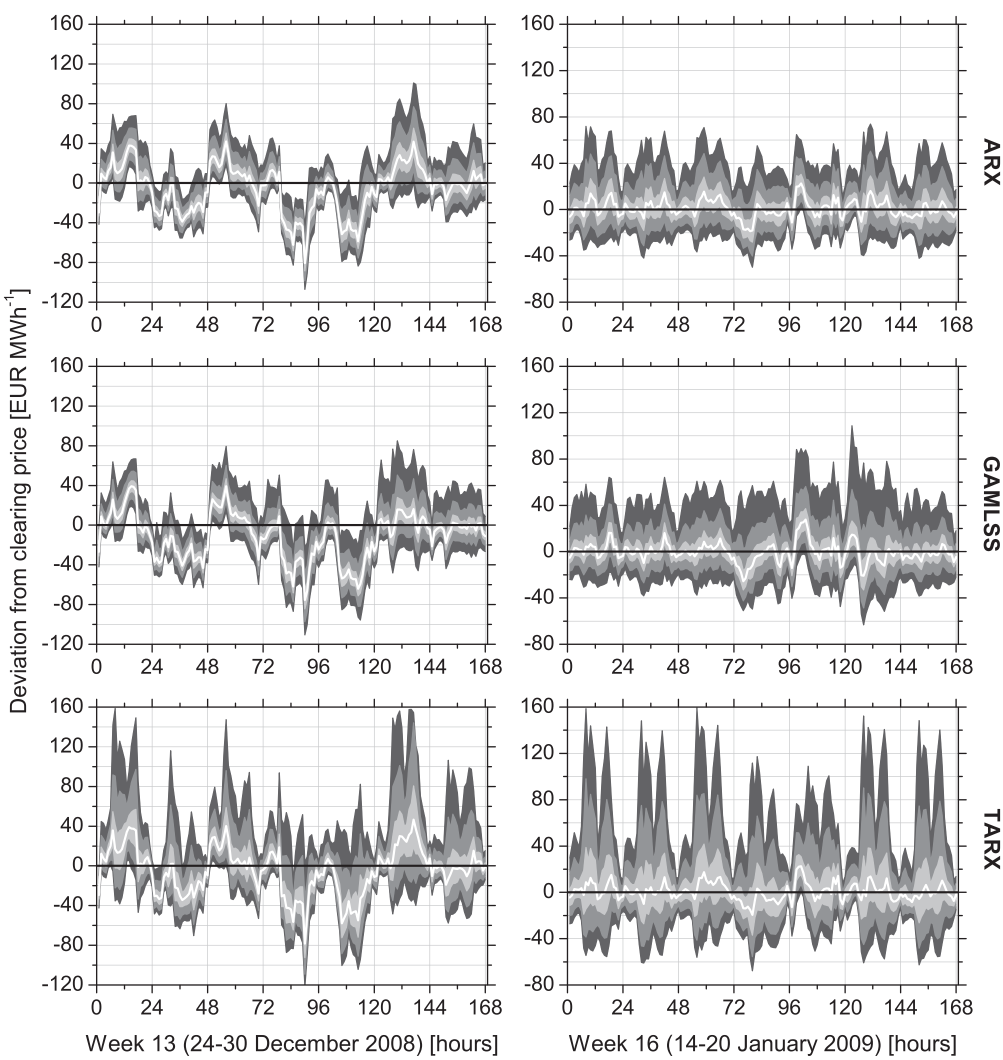

Fig. 5. Deviations of the point forecasts and limits of the 50%, 90%, and 99% two-sided CIs from the market clearing price for the 13th and 16th week of the testing period of IPEX data.

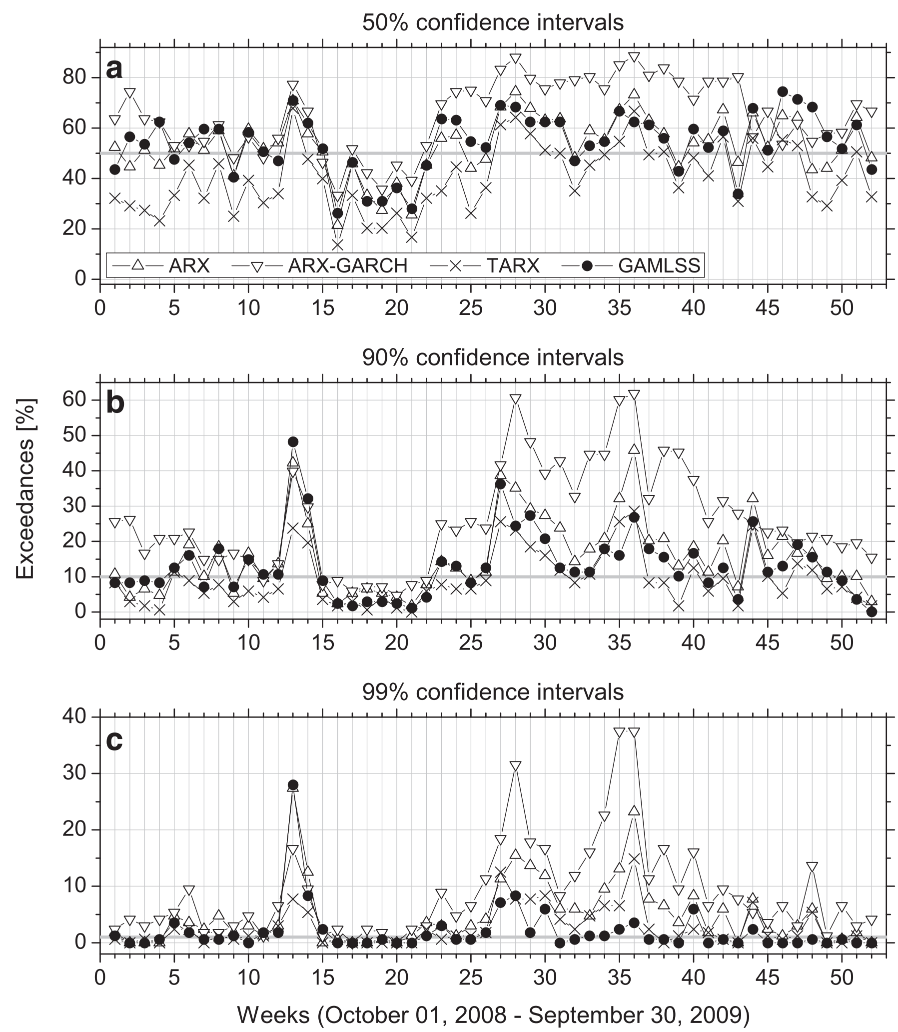

Fig. 6. Percentage of data falling outside the CIs with nominal coverage probability equal to (a) 50%, (b) 90%, and (c) 99% for each week of the testing period of IPEX data.

carry out one-day ahead forecasting of electricity market prices. While traditional approaches for time series analysis try to model the expected price pattern and the residuals, GAMLSS consider the prices as realizations of a variable whose distribution function parameters vary dynamically with time and other explanatory variables. Therefore, the complete instantaneous distribution function of price is available without performing separate analyses of the expected price and model residuals. Since the distribution parameters are linked to the statistical moments, and complex dependence functions between parameters and explanatory variables can be introduced in the modeling, GAMLSS allows accounting for nonstationarity of high order moments, say >2, in a coherent and flexible framework.

Nevertheless, this approach involves some subjectivity, common to all stochastic methods, which can influence the results. For example, even though the choice of the most appropriate distribution, link and dependence functions, and explanatory variables can be carried out through performance indices related to the maximum likelihood, such as AIC and SBC, the model structure has to be refined by a trial-and-error approach based on suitable performance measures according to forecasting goal. As shown in the applications, in some cases GAMLSS can yield results better than the other reference models; however, all methods can be improved by modifying the model setup appropriately. The advantage of GAMLSS is that these setup changes can be applied to expectation, variance, skewness, and kurtosis of price distribution within quite a simple framework similar to Generalized Linear Models and Generalized Additive Models. Moreover, since the comparison of different methods is fundamental to assess the reliability of the results, GAMLSS provide a further tool for the cross-validation of other forecasting procedures.

Supplementary materials related to this article can be found online at doi:10.1016/j.eneco.2011.05.001.

Acknowledgments

The author thanks two anonymous reviewers for their constructive comments. The usual disclaimer applies.

References

Aggarwal, S.K., Saini, L.M., Kumar, A., 2009. Electricity price forecasting in deregulated markets: a review and evaluation. International Journal of Electrical Power & Energy Systems 31, 13–22.
Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19 (6), 716–723.
Bessec, M., Bouabdallah, O., 2005. What causes the forecasting failure of Markov-switching models? A Monte Carlo study. Studies in Nonlinear Dynamics & Econometrics 9 (2) Article 6.
Blumsack, S., Perekhodtsev, D., Lave, L.B., 2002. Market power in deregulated wholesale electricity markets: issues in measurement and the cost of mitigation. The Electricity Journal 15 (9), 11–24.

Blumsack, S.A., Apt, J., Lave, L.B., 2006. Lessons from the failure of U.S. electricity restructuring. The Electricity Journal 19 (2), 15–32.
Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.
Bosco, B.P., Parisio, L.P., Pelagatti, M.M., 2007. Deregulated wholesale electricity prices in Italy: an empirical analysis. International Advances in Economic Research 13, 415–432.
Bunn, D.W., 2000. Forecasting loads and prices in competitive power markets. Proceedings of the IEEE 88 (2), 163–169.
Dacco, R., Satchell, S., 1999. Why do regime-switching models forecast so badly? Journal of Forecasting 18 (1), 1–16.
Fahrmeir, L., Lang, S., 2001. Bayesian inference for generalized additive mixed models based on Markov random field priors. Applied Statistics 50, 201–220.
Haldrup, N., Nielsen, M.Ø., 2006. A regime switching long memory model for electricity prices. Journal of Econometrics 135 (1–2), 349–376.
Hastie, T.J., Tibshirani, R.J., 1990. Generalized additive models. Chapman and Hall, London.
Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. International Journal of Forecasting 22, 679–688.
Johnson, N.L., 1949. Systems of frequency curves generated by methods of translation. Biometrika 36 (1–2), 149–176.
Kosater, P., Mosler, K., 2006. Can Markov regime-switching models improve power-price forecasts? Evidence from German daily power prices. Applied Energy 83 (9), 943–958.
Makridakis, S.G., Wheelwright, S.C., Hyndman, R.J., 1998. Forecasting: Methods and Applications, 3rd Edition. Wiley, New York.
McCullagh, P., Nelder, J.A., 1989. Generalized Linear Models, 2nd Edition. Chapman and Hall/CRC Press, London.
McCulloch, C.E., 1997. Maximum likelihood algorithms for generalized linear mixed models. Journal of the American Statistical Association 92 (437), 162–170.
McCulloch, C.E., 2003. Generalized Linear Mixed Models. IMS, Hayward, CA.
Misiorek, A., Weron, R., 2006. Interval forecasting of spot electricity prices. Proceedings of the International Conference "The European Electricity Market EEM-06", May 24–26, 2006, Warsaw, Poland.
Misiorek, A., Trueck, S., Weron, R., 2006. Point and interval forecasting of spot electricity prices: linear vs. non-linear time series models. Studies in Nonlinear Dynamics & Econometrics 10 Article 2.
Nelder, J.A., Wedderburn, R.W.M., 1972. Generalized linear models. Journal of the Royal Statistical Society Series A 135, 370–384.
Nogales, F.J., Conejo, A.J., 2006. Electricity price forecasting through transfer function models. Journal of the Operational Research Society 57, 350–356.
R Development Core Team, 2009. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria3-900051-07-0. URL http://www.R-project.org.
Rigby, R.A., Stasinopoulos, D.M., 2005. Generalized additive models for location, scale and shape. Applied Statistics 54, 507–554.
Schwarz, G., 1978. Estimating the dimension of a model. Annals of Statistics 6 (2), 461–464.
Serati, M., Manera, M., Plotegher, M., 2008. Modelling electricity prices: from the state of the art to a draft of a new proposal. Liuc papers in economics. Cattaneo University (LIUC. URL http://econpapers.repec.org/RePEc:liu:liucec:210.
Stasinopoulos, D.M., Rigby, R.A., 2007. Generalized additive models for location scale and shape (GAMLSS) in R. Journal of Statistical Software 23 (7).
Stasinopoulos, D.M., Rigby, R.A., Akantziliotou, C., 2007. gamlss: Generalized Additive Models for Location Scale and Shape. R package version 1.6-0. URL http://www.londonmet.ac.uk/gamlss/.
Villarini, G., Serinaldi, F., Smith, J.A., Krajewski, W.F., 2009a. On the stationarity of annual flood peaks in the continental United States during the 20th century. Water Resources Research 45, W08417.
Villarini, G., Smith, J.A., Serinaldi, F., Bales, J., Bates, P.D., Krajewski, W.F., 2009b. Flood frequency analysis for nonstationary annual peak records in an urban drainage basin. Advances in Water Resources 32 (8), 1255–1266.
Weron, R., 2006. Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach. Wiley, Chichester.
Weron, R., Misiorek, A., 2006. Short-term electricity price forecasting with time series models: a review and evaluation. In: Mielczarski, W. (Ed.), Complex Electricity Markets. Łódź, Poland, pp. 231–254. Ch.10.
Weron, R., Misiorek, A., 2008. Forecasting spot electricity prices: a comparison of parametric and semiparametric time series models. International Journal of Forecasting 24, 744–763.