Energy Economics 47 (2015) 98–111

Contents lists available at ScienceDirect

Energy Economics

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

Efficient modeling and forecasting of electricity spot prices

Florian Ziel *, Rick Steinert **, Sven Husmann

Europa-Universität Viadrina, Große Scharrnstraße 59, 15230 Frankfurt (Oder), Germany

ARTICLE INFO

Article history:
Received 6 May 2014
Received in revised form 6 October 2014
Accepted 7 October 2014
Available online 23 October 2014

JEL classification:
G17
C32
C51
C53
C58
Q41

Keywords:
Electricity price
Renewable energy
Leverage effect
Forecasting
VAR
Lasso

ABSTRACT

The increasing importance of renewable energy, especially solar and wind power, has led to new forces in the formation of electricity prices. Hence, this paper introduces an econometric model for the hourly time series of electricity prices of the European Power Exchange (EPEX) which incorporates specific features like renewable energy. The model consists of several sophisticated and established approaches and can be regarded as a periodic VAR-TARCH with wind power, solar power, and load as influences on the time series. It is able to map the distinct and well-known features of electricity prices in Germany. An efficient iteratively reweighted lasso approach is used for the estimation. Moreover, it is shown that several existing models are outperformed by the procedure developed in this paper.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

With the ongoing liberalization of electricity markets over the past decades, the volume of electricity traded via exchanges has greatly increased. This in turn led to an increasing transparency of the price for electricity. Due to the substantial dependence of companies and private households on this price, modeling electricity prices has become one of the cornerstones of research into the energy markets. But such modeling turns out to be at the edge of many disciplines in research. For instance, the analysis of the underlying trade mechanisms can be allocated to economics. But the energy production itself is a process which can be related to engineering and the rules governing the exchange of energy, especially renewable energy production, are determined by law and politics. Hence, modeling electricity prices can be a complex issue. This is also reflected in the time series, where many unusual but already stylized facts can be observed.

The model proposed in our paper tackles this complexity in several ways. Its distinctive features compared to the existing literature can be boiled down to seven key facts. Our approach: 1. Models the electricity price without any data manipulation, 2. Incorporates every established stylized fact of electricity prices, 3. Provides insights for the structure of the leverage effect, 4. Proves the effect of wind and solar power on price, 5. Accounts for specific holiday effects and daylight saving time effects in the wind and solar generation, 6. Does not need any future information to provide accurate forecasts. Finally, 7, it uses efficient and rapid state of the art estimation techniques.

We will fit our proposed model to the hourly electricity price of the European Power Exchange (EPEX) for the period of 28.09.2010 up to 01.05.2014. The data sets are obtained from the EPEX at www.epexspot.com for the hourly day-ahead spot price data of Germany/Austria, from the European Network of Transmission System Operators for Electricity at www.entsoe.eu for the hourly load data of Germany, and from the Transparency Platform of the European Energy Exchange (EEX) at www.transparency.eex.com for the hourly wind and solar power feed-in for Germany.[1]

Our paper is organized as follows. In Section 2 we give an overview about the setting in which electricity exchange takes place and name the distinct challenges occurring in modeling the time series. The

* Corresponding author. Tel.: +49 335 5534 2986.
** Corresponding author. Tel.: +49 335 5534 2988.
E-mail addresses: ziel@europa-uni.de (F. Ziel), steinert@europa-uni.de (R. Steinert), husmann@europa-uni.de (S. Husmann).

[1] We remark that the load and renewable data for Austria is not included in the data as our main focus is the German energy market.

http://dx.doi.org/10.1016/j.eneco.2014.10.012
0140-9883/© 2014 Elsevier B.V. All rights reserved.

subsequent section sets up our model, which aims to face those challenges. Section 4 presents the estimation procedures for our model. In Sections 5 and 6 we fit the model to the hourly electricity price time series of the European Power Exchange and apply a comprehensive forecast study. The last section concludes by discussing our findings.

2. Challenges in modeling electricity prices

Energy markets are a rapidly changing field of the economy. The liberalization of these markets and the subsequent development of the energy mix account for that fact. But as different countries worldwide face different preconditions, for instance politically or climatically, their energy markets tend to have a very heterogeneous structure. Hence, the findings for one country may not, or may only partly, be used for another country or region.

In the case of the German energy market, a substantial amount of the daily demand is traded via an exchange. Spot market trading takes place by continuous trading and auctions. Prices for this market can be obtained either from intraday trading or from day-ahead auction prices. The latter is represented by the EPEX spot auction price for Austria and Germany. EPEX is a member of the EEX group. Since 2008, the EEX has not prohibited negative prices (Keles et al., 2012). By considering all products of the exchange, there are currently 250 market participants at the EEX. Considering only the Austrian/German day-ahead auction of the EPEX results in 197 participating traders.

The non-negligible amount of market participants is also an aftermath of liberal energy laws in Germany, especially when renewable energy is considered. The Erneuerbare-Energien-Gesetz (EEG) and its corresponding enactments are governmental regulations which embody this idea of liberalization. As a consequence, the supply of energy from renewables rose significantly within the past years. Incentives, like feed-in-tariffs for renewables granted by the government, catalyzed this development. But the growing amount of power plants for renewables had a direct influence on the price for electricity, (Edenhofer et al., 2013), as such energy can be produced at a cost of almost zero (Würzburg et al., 2013).

Changes in the EEG in 2012 had direct impact on the marketing of renewable energy.[2] The EEG allows for marketing the produced renewable energy not only to TSOs or other market players but also directly at the exchange. According to §33g EEG, producers of such energy receive a market premium in addition to the market price for selling their electricity. Hence, also producers of renewable energy have an incentive to directly trade at the electricity exchange or to sell it via OTC business.

Nevertheless, other regulations have also affected the market directly. For instance, transmission system operators (TSO) are obliged to sell their electricity only on the day-ahead or intraday spot market, if they decide to use an exchange.[3] Hence, data since the publication of this regulation in 2009 are of special interest.

In the economic theory of competitive markets, the price of electricity should equal its marginal cost. As the feed-in of renewable energy with zero marginal cost replaces every other energy source with higher marginal cost, the price of electricity should decline (Keles et al., 2013). This is known as the merit-order effect. Furthermore, the demand curve for electricity can be assumed to be inelastic (Sensfuß et al., 2008) as a certain amount of power is needed regardless of the price. This implies that modeling the impact of renewables on the price of electricity is highly necessary, as those energy sources will always lead to a modified marginal cost structure as the share of renewable energy sources is increasing.

Empirical evidence for the reduction of electricity prices caused by the emergence of renewable energy has been shown by many authors

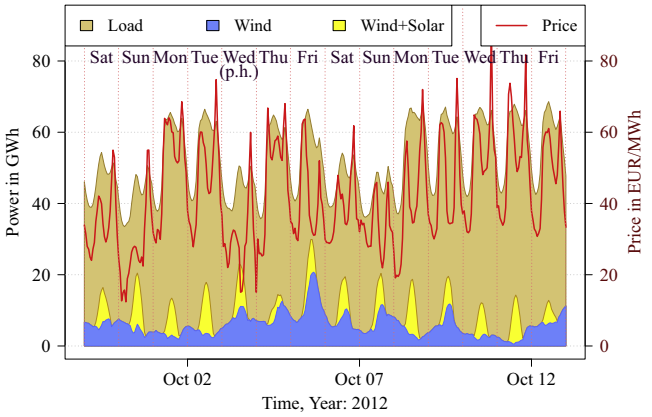

Fig. 1. The hourly load, wind and solar power feed-in with the corresponding spot price. The 3rd of October is a public holiday (German unity).

in the recent literature. For instance, Woo et al. (2011) use a regression analysis for the Texas electricity price market to examine the effect of wind power generation. Another multivariate regression approach was applied to German and Austrian electricity prices by Würzburg et al. (2013), where, among other time series, wind and solar power were examined and also led to a reduction in price. Huisman et al. (2013) obtained equivalent results for the Nord Pool market by modeling energy supply and demand.

This relation is also illustrated in Fig. 1, which shows the price, load, and the patterns of solar power and wind power for two weeks of October 2012. Whenever the combined effect of wind and solar power is high, the line representing the price seems to decline considerably.

But the introduction of renewables not only leads to a price reduction effect, it can also increase the goodness-of-fit for modeling the time series of prices. Concerning this, a comprehensive study was done by Cruz et al. (2011). They combined sophisticated models like Holt–Winters, ARIMA, and neural networks, with, e.g., wind power, and provided evidence for this inclusion's being beneficial to the price modeling. Also Yan and Chowdhury (2013), Liebl (2013) and Kristiansen (2012) obtained competitive goodness-of-fit statistics by including wind power in their approaches. Using a simulation study for the dependencies of the EEX electricity price and the wind power generation Keles et al. (2013) were able to show the advantages of including wind power.

The price of electricity is also heavily dependent on the day of the week and on the four seasons of the year (Weron, 2006). The reason for this is twofold. First, especially solar energy production depends on the period of sunshine: reduced, for instance, in winter. Second, the daily demand for electricity is dependent on the working days, e.g. whether industrial machines are running and require energy or not. An example of two observed months for illustrative purposes is given in Fig. 1. The light-brown shaded area represents the electric load pattern of the first two weeks of October in 2012. Comparing the price time series with the load time series indicates that both variables are positively related. As on weekends the load is usually lowered, weekends and weekdays[4] within the chosen period tend to exhibit different price patterns. This is also true for holidays, e.g. German unity day on the 3rd of October, where the price and the load are both lower than on other weekdays. A more detailed investigation of special days and phases of the day is done in Section 3.

The consideration of these effects is also an important topic in the literature. The weekly dependence is usually modeled in time series analysis by incorporating the equivalent lagged value, e.g. lag 168 for hourly data (e.g. in Kristiansen, 2012). Nevertheless, the consideration of

[2] Even more recent changes of the EEG and related laws in July and August 2014 are not considered within this article, as our dataset ends in May 2014.

[3] According to §2 AusglMechV.

[4] We use the term weekday to refer to the days from Monday to Friday.

special days or special phases of the day is done only rarely. Cancelo et al. (2008) use a combined ARIMA model to forecast load and provide evidence for the inclusion of special days within the estimation. Guthrie and Videbeck (2007) calculate a periodic autoregression for the half-hourly electricity price of the New Zealand Electricity Market. They divide the day into five different time periods and show that they differ from each other.

Another appealing approach is to let the coefficients of the proposed model vary over time. For instance, Karakatsani and Bunn (2008) present a comparative analysis of models with and without time varying parameters. They show for British half-hourly electricity prices that models with time varying parameters dominate other autoregressive approaches with constant parameters. A combined model for the inclusion of holidays and time-varying coefficients was introduced by Koopman et al. (2007). They use a Reg-ARFIMA-GARCH model for the EEX, Powernext and AAX price data.

Electricity is also distinct from most other commodities as it is not efficiently storable with the existing technology. This, therefore, adds another source of risk (Knittel and Roberts, 2005). Nevertheless, to a certain extent it can be regarded as indirectly storable because some energy sources such as fuel or gas can be stored and therefore used equivalently to fulfill obligations from, e.g., derivatives. (Huisman and Kilic, 2012).

The combination of almost non-storability, inelastic demand, and the strong dependence on highly fluctuating energy supplies causes the price of electricity to be endowed with unique characteristics. These are known as the stylized facts of electricity prices.

Most authors refer to three or even more characteristics. The most frequently mentioned are seasonality, mean-reversion, and high heteroscedastic volatility with extreme price spikes (e.g. Weron, 2006); Eydeland and Wolyniec, 2003). For illustrative purposes, the EPEX electricity price of the first two months of the year 2012 is depicted in Fig. 2. The pattern shows high price spikes in both directions, which usually last for several hours. After the impact of such a shock, the price reverts to its usual daily level.

Empirical consideration of negative prices in the German/Austrian electricity market is quite rare as they were not present until 2009. Moreover, modeling negative prices is also a difficult task, as some fitting and transformation approaches like logarithmic transformation are not possible. Hence, some authors simply cut off or shift the time series. Nevertheless, Keles et al. (2012) showed via simulation study for the EEX that the inclusion of negative prices yielded better results for their investigated models. Also Fanone et al. (2013) argue for the inclusion of negative prices.

Price spikes can be modeled in different ways. According to Christensen et al. (2012) there are mainly three different common approaches. These are specific autoregressive, Markov-regime-switching, and jump diffusion models. Other authors take advantage of a combination of these models, e.g. Bordignon et al. (2013) and Escribano et al. (2011). Besides these obvious characteristics of electricity prices, the recent literature has argued for an appreciation of the leverage effect, which is well-known in financial economics. A leverage effect accounts for the variance of a time series reacting asymmetrically to negative and positive price shocks (Black, 1976). The standard leverage effect has higher variance for negative than for positive price shocks. By analyzing hourly California electricity prices Knittel and Roberts (2005) were able to detect an inverse leverage effect for the time series. Their paper showed that the volatility responses were more intense to positive than to negative shocks. This was later acknowledged for instance by Bowden and Payne (2008) and Liu and Shi (2013). Cifter (2013), however, provided evidence for the standard leverage effect by examining the daily returns of the Nord Pool electricity price market.

As a matter of fact, modeling electricity prices is a complex issue. In addition to classic autoregression approaches, many researchers have applied methods from related scientific fields, e.g. heuristic or a combination of autoregression and other techniques. As a detailed overview of all those approaches would be outside the scope of this paper, the interested reader is referred to the works of Weron (2006) or Aggarwal et al. (2009) who provide a comprehensive overview of the recent literature.

In the following section we will address the challenges which occur when electricity prices are examined. We introduce our own model, which is tailor-made for the hourly EPEX electricity price data. It will be shown in detail how the challenges influence our approach and how they are tackled. We will refer to this model as periodic VAR-TARCH. It includes hourly data of the electricity load, and of wind and solar power feed-in. The model is capable of forecasting the price itself and also its random dependent variables.

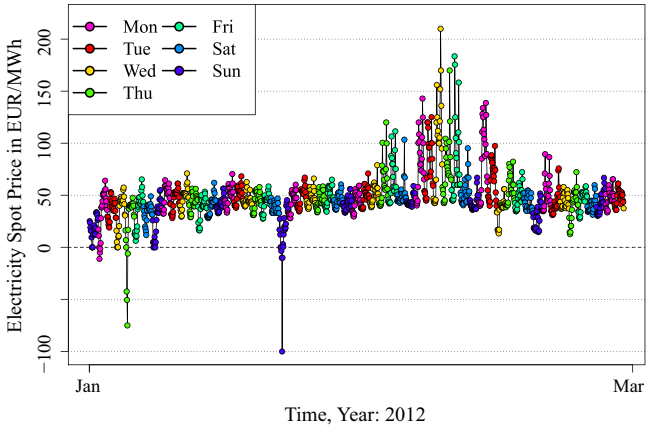

Fig. 2. Hourly electricity spot price from the 1st of January to the 1st of March 2012.

3. Model for electricity prices

Let $\mathbf{Y}$ be the considered multivariate process with dimension $d = 3$. So $\mathbf{Y}_t = \left(Y_t^{\mathcal{P}}, Y_t^{\mathcal{L}}, Y_t^{\mathcal{R}}\right)$ is a vector of the price of electricity, the load, and the renewable power feed-in at time t. Note that $Y_t^{\mathcal{R}}$ is the sum of the wind and solar power feed-in. Let $\mathcal{I} = \{\mathcal{P}, \mathcal{L}, \mathcal{R}\}$ be the corresponding index set for the three univariate processes.

The basic model is almost equivalent to a simple vector autoregressive model. It is given by the autoregressive model

$$Y_t^i = \mu^i(t) + \sum_{j\in\mathcal{I}} \sum_{k\in I_{i,j}} \phi_k^{i,j}(t) Y_{t-k}^j + \varepsilon_t^i \tag{1}$$

for $i \in \mathcal{I}$, where $\boldsymbol{\mu}(t) = \left(\mu^{\mathcal{P}}(t), \mu^{\mathcal{L}}(t), \mu^{\mathcal{R}}(t)\right)$ is a trend component, $\varepsilon_t = \left(\varepsilon_t^{\mathcal{P}}, \varepsilon_t^{\mathcal{L}}, \varepsilon_t^{\mathcal{R}}\right)$ are noise terms with zero mean and time dependent covariance matrix Σ_t, and they are assumed to be independent. Note that the autoregressive parameters $\phi_k^{i,j}(t)$ may also depend on time, such as the trend $\boldsymbol{\mu}(t)$ and the matrix Σ_t. The considered lags are given by the index sets $I_{i,j}$. These index sets are crucial and determine the autoregressive dependency structure within $\mathbf{Y}_t$. Here it is important to know that the estimation method described below contains an automatic model selection procedure. Hence, there is always a trade-off in the pre-selection of index sets. Giving the model large index sets $I_{i,j}$ can lead to better models as there is more possible information provided. But on the other side this increases the computation time and the probability of modeling noise. Thus, especially the index sets $I_{i,j}$ have to be chosen carefully.

Table 1
Considered lags of the index sets $I_{i,j}$ for $i, j \in \mathcal{I}$.

Index sets	Contained lags
$I_{\mathcal{P},\mathcal{P}}, I_{\mathcal{L},\mathcal{L}}, I_{\mathcal{P},\mathcal{L}}$	1, 2, ..., 361, 504, 505, 672, 673, 840, 841, 1008, 1009
$I_{\mathcal{P},\mathcal{R}}, I_{\mathcal{L},\mathcal{R}}$	1, 2, ..., 49
$I_{\mathcal{R},\mathcal{P}}, I_{\mathcal{R},\mathcal{L}}, I_{\mathcal{L},\mathcal{P}}$	–
$I_{\mathcal{R},\mathcal{R}}$	1, 2, ..., 361

The index sets $I_{i,j}$ we used are given in Table 1. On first view the chosen lags in $I_{i,j}$ seem more or less random, albeit every lag is chosen by employing statistical facts of the data.

For most index sets $I_{i,j}$ the first lags 1, 2, … 361 simply model the linear dependence of the past 361 h, so 2 weeks plus a day and an hour. For $I_{\mathcal{P},\mathcal{R}}$ and $I_{\mathcal{L},\mathcal{R}}$ we chose a dependence of the past 49 h. But we also added some larger lags like 504 and 505 to $I_{\mathcal{P},\mathcal{P}}$, $I_{\mathcal{L},\mathcal{L}}$, and $I_{\mathcal{P},\mathcal{L}}$ which account for a possible dependence on the price of the past weeks.

These lags allow a structure in the process that is similar to a multiplicative seasonal AR model. For example, the multiplicative structure of the lag polynomials in a seasonal AR(1) × $(2)_{168}$ is given by $(1 - a_1B)(1 - a_{168}B^{168} - a_{336}B^{336}) = 1 - a_1B - a_{168}B^{168} + a_1a_{168}B^{169} + a_{336}B^{336} + a_1a_{336}B^{337}$ with B as backshift or lag operator.

Thus it contains lags 1, 168, 169, 336, and 337. Using this evaluation of the lag-polynomial for higher order seasonal AR processes, we receive that lags like 504, 505, 672, 673, 840, 841, 1008 or 1009 should be contained in the model. Such seasonal lags were also considered by different authors in the past, such as Liu and Shi (2013).

As mentioned above, we could also increase index sets, e.g. replacing $I_{\mathcal{P},\mathcal{P}}$ by {1, …, 1009}. However, we noticed during our study that many parameters between $(k - 1)168 + 1$ and $k168$ are zero for large k, i.e. $k \geq 3$. Another simple but useful approach for selecting appropriate index sets is based on evaluating the sample partial autocorrelation function of $\mathbf{Y}_t$; the lags that correspond to large absolute values should be included. Indeed this approach leads to very similar index sets compared to ours. It is worth mentioning that we observe large PACF values at lags 504 and 505 which coincides with our discussion above. The preselection of lags therefore provides some flexibility to the applicant, as lags, which are just assumed to be significant can be implemented very easily and will still undergo a statistical reliable lag selection process.

Note that we choose $I_{\mathcal{P},\mathcal{R}}$ and $I_{\mathcal{L},\mathcal{R}}$ as being small, as we assume that there is no impact of renewable energy feed-in to the price to last for more than two days. Moreover, the renewable energy feed-in does not have any weekly seasonal structure, only a daily and a yearly one. Thus, $I_{\mathcal{R},\mathcal{R}}$ does not contain higher order weekly based lags like 504 and 505 as e.g. $I_{\mathcal{P},\mathcal{P}}$.

The trend component $\boldsymbol{\mu}(t)$ is modeled by a linear combination of $M(\mu^i)$ basis functions plus a linear trend, so we have

$$\mu^i(t) = \mu_0^i + \mu_{\text{lin.}}^i t + \sum_{j=1}^{M(\mu^i)} \mu_j^i \widetilde{B}_j^{\mu^i}(t) \tag{2}$$

for $i \in \mathcal{I}$ with parameters μ_j^i and basis functions $\widetilde{B}_j^{\mu^i}(t)$. The structure of the used basis functions is inspired by periodic basis functions. This enables modeling recurring events like specific days of the week. Nevertheless, as some events like holidays are not periodic, we are allowing $\widetilde{B}_j^{\mu^i}(t)$ to be not strictly periodic. As basic concept for the construction of $\widetilde{B}_j^{\mu^i}(t)$ we consider periodic B-splines as used in Harvey and Koopman (1993).

However, this approach using local B-splines was mainly chosen to increase the flexibility and the interpretability of the model. We therefore want to point out, that other, e.g. Fourier approaches, might lead to estimation results of comparable quality.

The chosen B-splines are denoted as $B_{T,d_\mathcal{K}}(t) = B_{T,d_\mathcal{K}}(t; \mathcal{K}(d_\mathcal{K}, T, D), D)$, where D represents the degree of the spline and $\mathcal{K} = \mathcal{K}(d_\mathcal{K}, T, D) = \{k_0, \ldots, k_{D+1}\}$ the set of equidistant knots $k_0 < \cdots < k_{D+1}$, that have equal distance $d_\mathcal{K}$ and are centered around T. Hence, we have $k_0 = T - d_\mathcal{K}\frac{D+1}{2}$, $k_{D+1} = T + d_\mathcal{K}\frac{D+1}{2}$, and if we select an odd degree D we get $k_{\frac{D+1}{2}} = T$. The distance $d_\mathcal{K} = k_1 - k_0$ is important for applications as it accounts for the degree of approximation of the time dependent effects and affects the amounts of parameter in the model, the smaller $d_\mathcal{K}$ the better the corresponding approximation, but the larger the considered parameter space. For D we choose $D = 3$, which provides the popular cubic splines which are twice continuously differentiable.

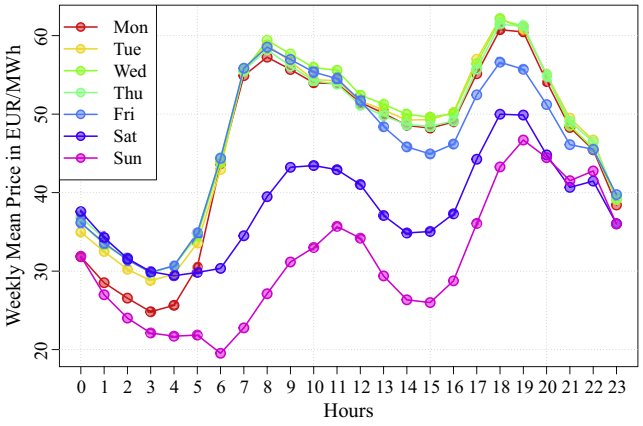

Fig. 3. Average price in EUR/MWh for every hour, separated by every day of the week.

For mirroring the movement of the electricity price data we want that $\mu^i(t)$ smoothly changes over the days and during the week as well. So, for example, the coefficient for Sunday at 8 am is different to the coefficients at Sunday 9 am or Monday 8 am, but it is equal for every Sunday at 8 am.

However, we also observe that the mean behavior of at least some days is very similar to that of other days, see Fig. 3. This finding is embedded in our model by proposing the assumption of some days having equal hourly mean components, which in turn reduces the number of parameters in the model. For example, the mean at Tuesday 8 am is assumed to be the same as on a Wednesday 8 am. Finally we consider five groups of parameter restrictions: Saturday, Sunday, Monday from 0 am to 12 am, Friday from 12 am to 12 pm and the core week from Monday 12 am to Friday 12 am. The Monday morning and Friday afternoon and evening have to be considered to get an appropriate phase-in and phase-out of the weekend. Friday, represented as the cyan colored line, departs from the typical structure of the other days (excluding the weekend) approximately at 12 am. It is almost the same for Monday, with the difference that Monday needs 7 to 8 h to reach an equivalent structure comparable to the other weekdays. In order to maintain equidistant phase-in and phase-out periods, we set both periods to 12 h.

Furthermore, there are significant impacts of public holidays, as mentioned by several authors. We consider every German public holiday as a Sunday. In addition to that there are some regional public holidays in Germany which are celebrated only by some areas. Every regional public holiday that concerns at least 25% of the population and is not on a Sunday, is treated as a Saturday as well. Moreover, we consider Christmas Eve and New Years Eve as regional public holidays. Consequently we take the 12 h after a (regional) public holiday as Monday-am and the 12 h before every (regional) public holiday as Friday-pm. Of course, this applies only if the day itself is not a Sunday, Saturday or (regional) public holiday. An overview of the explained day-grouping is given in Table 2. In addition, the table provides the number of unique hours within each group. Unique refers to an hour having one unique parameter. For instance, in the case of the group *normal*, there are only 24 unique hours, even though its contained days, from Monday lunch time to Friday lunch time, have 96 h in total. This approach accounts for the mean of hours of different days within this group being very similar.

To employ our model we denote the corresponding five groups of Table 2 as $\mathbb{G} = \{1, \ldots, 5\}$ and create a time set $\mathbb{T}_g$. Every group g is an element of $\mathbb{G}$, namely $g \in \mathbb{G}$. The combined time set $\mathbb{T}_g$ for all groups contains the integer value of the first hour of every group and day. The subset for the first group $\mathbb{T}_1$, for instance, contains only those integer values of hours, which lay inside the group *full off* and represent the

Table 2
Group structure of the days.

Group	Contained days	Hours with different parameters
Full off	Sundays and public holidays	24
Semi off	Saturdays and regional public holiday if not *full off*	24
Phase in	12 h before *full* or *semi off* if not *full* or *semi off*	12
Phase out	12 h after *full* or *semi off* if not *full* or *semi off*	12
Normal	Others	24

hour 0 am of this day. For the group *phase in* and *normal* however, the subsets $\mathbb{T}_3$ and $\mathbb{T}_5$ contain the values of the hour 12 am for each day. The differences of the elements of $\mathbb{T}_g$ are therefore usually 12, 24 or 168, but can also be a different number, when a public holiday occurs. An example for such a set $\mathbb{T}_g$ could therefore be $\mathbb{T}_1 = \{\ldots, 0, 168, 336, 504, 528, 672, \ldots\}$, where at hour 528 (Monday) a holiday occurred.

Within each group g we set a specific number of basis functions. This number is related to the unique hours every group has. We determine the vector $N_{\mathbb{G}}$ as $N_{\mathbb{G}} = \left(d_{\mathcal{K}}^{\text{weekly}}\right)^{-1}(24, 24, 12, 12, 24)'$. Please notice that we divide the amount of unique hours by $d_{\mathcal{K}}^{\text{weekly}} = 4$, so that for instance the group *full off* contains $24/4 = 6$ different basis functions. In order to label these basis functions distinctively, we introduce the vector $C_{\mathbb{G}} = \left(d_{\mathcal{K}}^{\text{weekly}}\right)^{-1}\left(0, 24, 48, 60, 72, 96 - d_{\mathcal{K}}^{\text{weekly}}\right)'$, which contains the cumulative sum of the hours in $N_{\mathbb{G}}$ as entries. The subtraction of $d_{\mathcal{K}}^{\text{weekly}}$ in the last element of $C_{\mathbb{G}}$ is to avoid the singularity issue that occurs as the sum of all basis functions is constant. Then we define the function $G(\iota, C_{\mathbb{G}}) = g$ for $C_{\mathbb{G}}(g) < \iota \leq C_{\mathbb{G}}(g+1)$. The index variable ι is integer valued, starts from 1 and contains in our case every natural number up to $(96-4)/4 = 23$. Hence, the function $G(\iota, C_{\mathbb{G}})$ matches every ι, which represents the label of a basis function, to its specific group.

Moreover, we have to introduce the set $\mathcal{C} = \{C_{\mathbb{G}}(g) + 1 | g \in \mathbb{G}\}$ which will contain the index of the first basis function of each group. In our case this turns out to be $\mathcal{C} = \{1, 7, 13, 16, 19\}$. The resulting basis functions $\widetilde{B}_j^{\mu^i}(t)$ can then be constructed as follows:

$$\widetilde{B}_j^{\mu^i}(t) = \sum_{k \in \mathbb{T}_{G(j, C_{\mathbb{G}})}} B_{k, d_{\mathcal{K}}^{\text{weekly}}(t)}, \tag{3}$$

for $j \in \mathcal{C}$. The other basis functions can be obtained by shifting these basis functions. For $j \in \{C_{\mathbb{G}}(g)+2, \ldots, C_{\mathbb{G}}(g+1)\}$ they are iteratively defined as $\widetilde{B}_j^{\mu^i}(t) = \widetilde{B}_{j-1}^{\mu^i}\left(t - d_{\mathcal{K}}^{\text{weekly}}\right)$. These basis functions apply for the price and the load data, so $i \in \{\mathcal{P}, \mathscr{L}\}$. Fig. 4 displays a practical application of Eq. (3) to the time from 22.09.2012 to 05.10.2012. It shows the first

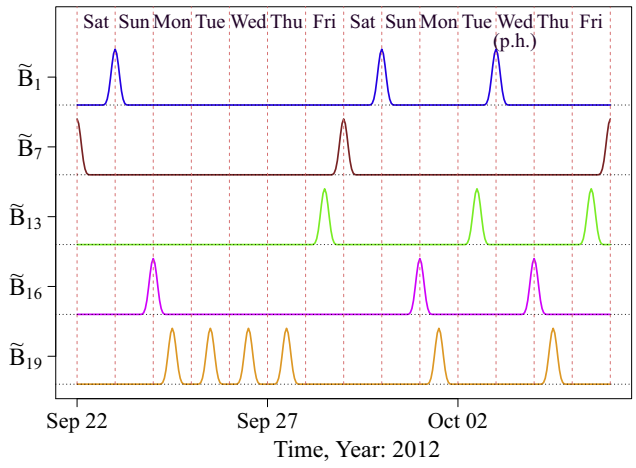

Fig. 4. First basis function $\widetilde{B}_j^{\mu^i}$ for $i \in \{\mathcal{P}, \mathscr{L}\}$ and each group $g \in \mathbb{G}$ from 22.09.2012 to 05.10.2012.

basis function of each of the five groups. The model automatically detects that the 03.10.2012 is a public holiday and therefore breaks the usual periodicity for some of the groups, whereas the other groups remain stable. The group *normal* for instance leaves out the basis functions for Tuesday and Wednesday, even though they would usually be considered within this group. However, the group *full off* does now include the Wednesday as public holiday and therefore gets a basis function for this day.

Furthermore, we can observe an annual pattern within the data, most distinctively for the load and solar data. Hence, we added B-splines with a yearly periodicity of $S_{\text{annual}} = 24 \times 365.24 = 8765.76$. As we assume that the meteorological processes have a smoother change over the year we added only $\eta_{\mathscr{R}}^{\text{annual}} = 6$ basis functions, whereas we take $\eta_i^{\text{annual}} = 12$ basis functions for the price and load, so $i \in \{\mathcal{P}, \mathscr{L}\}$. This leads to $d_{\mathcal{K},\mathscr{R}}^{\text{annual}} = 1460.96$ and $d_{\mathcal{K},i}^{\text{annual}} = 730.48$. For the wind and solar components we assumed no weekly structure, only a daily and a yearly one. Therefore we construct basis function for the daily structure with a periodicity of $S_{\text{daily}} = 24$ and annual basis functions. We choose $d_{\mathcal{K}}^{\text{daily}} = 4$, thus there are $\eta^{\text{daily}} = 6$ basis functions added. We also take into account interactions between the yearly and daily components, i.e. how the daily cycle behavior changes over the year. This seems important as, for instance, the length of the sunshine period of the day changes over the year and should have a strong influence. This is modeled by a multiplication of both components.

In detail the additional basis functions for the price and load are given by

$$\widetilde{B}_j^{\mu^i}(t) = \sum_{k \in \mathbb{Z}} B_{kS_{\text{annual}}, d_{\mathcal{K},i}^{\text{annual}}}(t)$$

for $j = \left(d_{\mathcal{K}}^{\text{weekly}}\right)^{-1} 96$ and $i \in \{\mathcal{P}, \mathscr{L}\}$. The other basis functions can be obtained by shifting these basis functions. For $j \in \left\{\left(d_{\mathcal{K}}^{\text{weekly}}\right)^{-1}96 + 1, \ldots, \left(d_{\mathcal{K}}^{\text{weekly}}\right)^{-1}96 + \eta_i^{\text{annual}} - 1\right\}$ they are iteratively defined as $\widetilde{B}_j^{\mu^i}(t) = \widetilde{B}_{j-1}^{\mu^i}\left(t - d_{\mathcal{K},i}^{\text{annual}}\right)$. Hence the number of basis functions $M\left(\mu^i\right) = \left(d_{\mathcal{K}}^{\text{weekly}}\right)^{-1}96 + \eta_i^{\text{annual}} - 1$ for $i \in \{\mathcal{P}, \mathscr{L}\}$.

For the wind and solar components we again define the basis functions for the daily and annual cycles as:

$$\widetilde{B}_{\text{daily},2}^{*,\mu^{\mathscr{R}}}(t) = \sum_{k \in \mathbb{Z}} B_{kS_{\text{daily}}, d_{\mathcal{K}}^{\text{daily}}}(t) \quad \text{and} \quad \widetilde{B}_{\text{annual},2}^{*,\mu^{\mathscr{R}}}(t) = \sum_{k \in \mathbb{Z}} B_{kS_{\text{annual}}, d_{\mathcal{K}}^{\text{annual}}}(t).$$

Please remark that we set $\widetilde{B}_{\text{daily},1}^{*,\mu^{\mathscr{R}}}(t) = \widetilde{B}_{\text{annual},1}^{*,\mu^{\mathscr{R}}}(t) = 1$ which will become necessary in a later step, where we model the interactions. The subsequent B-splines are again iteratively defined with $\widetilde{B}_{\text{daily},h_1}^{*,\mu^{\mathscr{R}}}(t) = \widetilde{B}_{\text{daily},h_1}^{*,\mu^{\mathscr{R}}}\left(t - d_{\mathcal{K}}^{\text{daily}}\right)$ for $h_1 \in \{3, \ldots, \eta^{\text{daily}}\}$ and $\widetilde{B}_{\text{annual},h_1}^{*,\mu^{\mathscr{R}}}(t) = \widetilde{B}_{\text{annual},h_1}^{*,\mu^R}\left(t - d_{\mathcal{K}}^{\text{annual}}\right)$ for $h_2 \in \{3, \ldots, \eta_{\mathscr{R}}^{\text{annual}}\}$. For modeling the interactions of the daily and annual cycles we can now introduce a modified version of basis functions for the renewables:

$$\widetilde{B}_j^{*,\mu^{\mathscr{R}}}(t) = \widetilde{B}_{\text{daily},h_1}^{*,\mu^{\mathscr{R}}}(t)\widetilde{B}_{\text{annual},h_2}^{*,\mu^{\mathscr{R}}}(t),$$

where $h_1 = j \bmod \eta^{\text{daily}} + 1$ and $h_2 = j \,\text{div}\, \eta_{\mathscr{R}}^{\text{annual}} + 1$ for $j \in \{1, \ldots, \eta^{\text{daily}}\eta_{\mathscr{R}}^{\text{annual}} - 1\}$. The term mod stands for the modulo operator and div is the integer without the remainder of a division. In this situation our previous definition of the first daily and annual basis functions is used, as the total set of B-spline now contains not only the simple daily and annual basis functions but also the interactions of them. Hence, we have included an amount of $M(\mu^{\mathscr{R}}) = \eta^{\text{daily}}\eta_{\mathscr{R}}^{\text{annual}} - 1$ basis functions.

Furthermore, in Eq. (1), the parameters $\phi_k^{i,j}(t)$ might also depend on time. So the autoregressive behavior can also change over time, especially in a seasonal manner. A related approach was also used by Koopman et al. (2007) and Bosco et al. (2007) for modeling daily electricity prices, and by Guthrie and Videbeck (2007) for intra-day prices. We assume that this dependence has a structure as in Eq. (2), so for $\phi_k^{i,j}(t)$ it is given by

$$\phi_k^{i,j}(t) = \phi_{k,0}^{i,j} + \sum_{j=1}^{M_i(\phi_k^{i,j})} \phi_{k,j} \widetilde{B}_j^{\phi_k^{i,j}}(t).$$

Again we have to choose the basis functions $\widetilde{B}_j^{\phi_k^{i,j}}(t)$, but from an application point of view it is more practical to take the same basis functions as for μ^i, which is implemented in our approach.

Unfortunately it would expand the parameter space enormously if we assume that for every possible lag $k \in I_{i,j}$ the coefficient $\phi_k^{i,j}(t)$ changes over time. Therefore, we let only those coefficients $L_{i,j} \subseteq I_{i,j}$ to change over time, which we assume to be most important. These are summarized in Table 3 and the other ones are empty.

Furthermore, there is a special impact on $\boldsymbol{Y}_t$ that has not been yet modeled. Due to daylight saving time, we have a time shift in March and October. As we removed the added hour in October and extrapolated the missing hour in March, there is no problem in modeling the electricity spot price as the population adjusts their every day behavior very quickly. The same holds true for the electricity load if we assume that the influence of regenerative power feed-in is negligible. But for the renewable energy feed-in this looks different, here we can improve our model. In winter the sun peak is at 12 am and in summer it is at 1 pm. This has an impact especially on the solar power. Fig. 5 illustrates this effect on the observed solar power feed-in in the morning. It can be observed that right after the time shift in March there is a downwards jump in the solar power feed-in present, and after the time shift in October a jump upwards occurs. We will model this behavior by a shift in the considered basis functions. Therefore, let $\mathbb{T}_{\text{DST}}$ denote the set of all time points with summer time. Then the considered time adjusted basis functions are given by

$$\widetilde{B}_j^{\mu^{\mathscr{R}}}(t) = 1_{\{t \notin \mathbb{T}_{\text{DST}}\}} \widetilde{B}_j^{*,\mu^{\mathscr{R}}}(t) + 1_{\{t \in \mathbb{T}_{\text{DST}}\}} \widetilde{B}_j^{*,\mu^{\mathscr{R}}}(t+1).$$

This shift is applied to all basis function of the wind and solar power feed-in. For the mean component $\mu^{\mathscr{R}}(t)$ the effect is clear, here this is equivalent to a shift in the hour labels. In contrast, the effect on the autoregressive parameters is not that obvious. The feed-in $Y_t^{\mathscr{R}}$ depends e.g. on $\phi_{24}^{\mathscr{R},\mathscr{R}}(t)$ before the time shift and is different to $\phi_{24}^{\mathscr{R},\mathscr{R}}(t)$ after the time shift. But $\phi_{24}^{\mathscr{R},\mathscr{R}}(t)$ would be the same as $\phi_{24}^{\mathscr{R},\mathscr{R}}(t-25)$ in March or $\phi_{24}^{\mathscr{R},\mathscr{R}}(t-23)$ in October, due to the time shift in daylight saving time. Nevertheless, in our approach this is not the case, as we explicitly model the daily and yearly interaction effects. The coefficients therefore truly embed the effects of the time shift.

But also the variance structure of the model is relevant. Here we assume that $\Sigma_t = \text{diag}(\sigma_t) = \text{diag}(\sigma_t^{\mathcal{P}}, \sigma_t^{\mathscr{L}}, \sigma_t^{\mathscr{R}})$. The error term ε_t can be written as

$$\varepsilon_t^i = \sigma_t^i Z_t^i \quad \text{where} \quad \left(Z_t^i\right)_{t\in\mathbb{Z}} \text{ is i.i.d. with } \mathbb{E}\left(Z_t^i\right) = 0 \text{ and } \mathbb{V}\text{ar}\left(Z_t^i\right) = 1.$$

As modeled by Keles et al. (2012) we want to explain conditional heteroscedasticity by an ARCH and GARCH type model. This approach

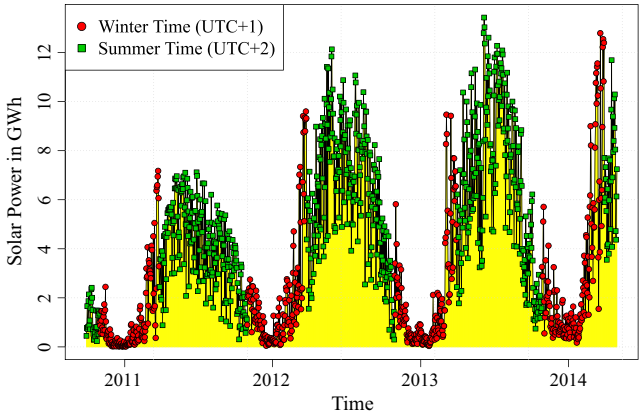

Fig. 5. Solar power feed-in at 8:00 am.

is suitable to model the price spikes, especially if the underlying distribution of Z_t is heavy tailed.

But as the price process $\mathbf{Y}_t$ is heavy tailed due to spikes, the residual process ε_t is likely to be heavy tailed with a small tail index, too. We noticed that it is not suitable to consider the square of ε_t^i, as done in a typical GARCH process, because they become even more heavy tailed. This halves the corresponding tail index and the convergence of the estimators will get worse. Instead of taking squares, we use the absolute value $|\varepsilon_t^i|$, which keeps the tail index on the same level. Some GARCH-type processes already use this structure, e.g. the TARCH model by Rabemananjara and Zakoïan (1993). Additionally, a TARCH and TGARCH model can handle the so called leverage effect, so that negative and positive past residuals have different influences on the volatility. We assume a TARCH process for σ_t so that

$$\sigma_t^i = \alpha_0^i(t) + \sum_{k\in J_i} \alpha_k^{+,i} \varepsilon_{t-k}^{+,i} + \alpha_k^{-,i} \varepsilon_{t-k}^{-,i} \tag{4}$$

holds with $\alpha_k^{+,i}, \alpha_k^{-,i} \geq 0$ where $\varepsilon_t^{+,i} = \max\{\varepsilon_t^i, 0\}$ and $\varepsilon_t^{-,i} = \max\{-\varepsilon_t^i, 0\}$. For the TARCH process we will assume that only the trend component $\alpha_0^i(t)$ varies seasonally over time. Thus the full parameter vector α_i is given by $\left(\alpha_{0,0}^i, \ldots, \alpha_{0,M(\alpha_0^i)}^i, \left(\alpha_k^{+,i}, \alpha_k^{-,i}\right)_{k\in J_i}\right)$. For $\alpha_0^i(t)$ we consider the same basis functions as for the corresponding $\mu^i(t)$, but without the linear trend, as it could conflict the positivity constraint to the parameters. Here we make use of another advantage of the periodic B-spline approach in contrast to a Fourier approximation. The positivity of the considered basis functions gives us automatically the restriction that the corresponding parameters have to be positive as well as σ_t must be positive.

The index sets J_i that we use are given in Table 4, they contain the typical discussed lags.

By multiplying $\gamma^i = \mathbb{E}|Z_t^i|$ and adding $\nu_t^i = \sigma_t^i(|Z_t^i| - \gamma^i)$ on both sides of Eq. (4) we get

$$|\varepsilon_t^i| = \widetilde{\alpha}_0^i(t) + \sum_{k\in J_i} \widetilde{\alpha}_k^{+,i} \varepsilon_{t-k}^{+,i} + \widetilde{\alpha}_k^{-,i} \varepsilon_{t-k}^{-,i} + \nu_t^i, \tag{5}$$

where $\widetilde{\alpha}_0^i(t) = \gamma^i \alpha_0^i(t), \widetilde{\alpha}_k^{+,i} = \gamma^i \alpha_k^{+,i}$, and $\widetilde{\alpha}_k^{-,i} = \gamma^i \alpha_k^{-,i}$. Here ν_t^i is a weak white noise process with $\mathbb{E}\left(\nu_t^i\right) = 0$. The auxiliary model in Eq. (5) is

Table 3
Considered lags of the index sets $L_{i,i}$ for $i \in \mathscr{T}$.

Index sets	Lags
$L_{\mathcal{P},\mathcal{P}}, L_{\mathscr{L},\mathscr{L}}$	1, 2, 24, 25, 168, 169
$L_{\mathscr{R},\mathscr{R}}$	1, 2, 23, 24, 25

Table 4
Considered lags of the index sets J_i with $i \in \mathscr{T}$ for the TARCH part.

Index sets	Lags
$J_{\mathcal{P}}, J_{\mathscr{L}}$	1:361, 504:505, 672:673, 840:841
$J_{\mathscr{R}}$	1:361

useful as it allows us to estimate α_i given estimates of ε^i. The fitted values $\widetilde{\sigma}_t^i$ of Eq. (5) are proportional to σ_t^i up to the constant γ^i. γ^i is the first absolute moment of ε_t^i which might help to characterize the distribution of ε_t^i. If ε_t^i follows a normal distribution it is exactly $\sqrt{2\pi^{-1}}\approx 0.798$, whereas, e.g., the standardized t-distributions have a larger first absolute moment.

Given reasonable estimates $\hat{\varepsilon}_t^i$ and $\widehat{\widetilde{\sigma}}_t^i$ for ε_t^i and $\widetilde{\sigma}_t^i$ we can use the plug-in estimator for γ^i to estimate it. With $\kappa_t^i = \left(\widehat{\widetilde{\sigma}}_t\right)^{-1}\hat{\varepsilon}_t^i$ this is given by

$$\hat{\gamma}^i = \left(\frac{1}{n-1}\sum_{t=1}^{n}\left(\kappa_t^i - \overline{\kappa^i}\right)^2\right)^{-\frac{1}{2}}, \tag{6}$$

as $\mathbb{V}\mathrm{ar}\left(\kappa_t^i\right) = \left(\gamma^i\right)^{-2}$. σ_t^i can now be estimated with $\hat{\sigma}_t^i = \left(\hat{\gamma}^i\right)^{-1}\widehat{\widetilde{\sigma}}_t^i$.

Finally, we can rewrite Eq. (1) in matrix notation with

$$\boldsymbol{Y}^i = \boldsymbol{X}^i\boldsymbol{\theta}_i + \varepsilon^i \tag{7}$$

with $\boldsymbol{Y}^i = (Y_1^i, \ldots, Y_n^i)$, $\mathbf{X}^i = (\mathbf{X}_1^i, \ldots, \mathbf{X}_n^i)$, $\varepsilon^i = (\varepsilon_1^i, \ldots, \varepsilon_n^i)$ and parameter vector $\boldsymbol{\theta}_i$ for $i \in \mathscr{I}$ given time points $t \in \{1, \ldots, n\}$. The vectors $\mathbf{X}_t^i$ are given through Eq. (2) by

$$\mathbf{X}_t^i = \left(1, t, \mathbf{B}_t\left(\mu^i\right), \mathbf{B}_t^{Y^i}\left(I_{i,\mathscr{P}}\right), \mathbf{B}_t^{Y^i}\left(I_{i,\mathscr{L}}\right), \mathbf{B}_t^{Y^i}\left(I_{i,\mathscr{R}}\right), \mathbf{B}_t^{Y^i,\mathrm{per}}\left(L_{i,i}\right)\right)$$

where we have $\mathbf{B}_t\left(\mu^i\right) = \left(\widetilde{B}_1^{\mu^i}(t), \ldots, \widetilde{B}_{M(\mu^i)}^{\mu^i}(t)\right)$, $\mathbf{B}_t^{Y^i}\left(I_{i,j}\right) = \left(Y_{t-k}^i\right)_{k\in I_{i,j}}$, for $j \in \mathscr{I}$, and $\mathbf{B}_t^{Y^i,\mathrm{per}}\left(L_{i,i}\right) = \left(Y_{t-k}^i\mathbf{B}_t\left(\mu^i\right)\right)_{k\in L_{i,i}}$. In the following we will refer the parameters that correspond to $(1, t, \mathbf{B}_t(\mu^i))$ as deterministic component, $\mathbf{B}_t^{Y^i}\left(I_{i,\mathscr{P}}\right)$ as price component, $\mathbf{B}_t^{Y^i}\left(I_{i,\mathscr{L}}\right)$ as load component, $\mathbf{B}_t^{Y^i}\left(I_{i,\mathscr{R}}\right)$ as wind and solar components and $\mathbf{B}_t^{Y^i,\mathrm{per}}\left(L_{i,i}\right)$ as periodic component given an $i \in \mathscr{I}$.

All in all our model has approximately 3500 possible parameters, which are given in Table 5. They match the mentioned deterministic, price, load, wind and solar, and periodic components defined above. Please notice that these parameters only represent the maximum amount of possible parameters, as our estimation procedure, which will be explained in the next section, will eliminate parameters, which are not significantly different from zero.

4. Estimation method

For the estimation of the parameters we will consider a variation of the iteratively weighted least squares approach as used for example in Mbamalu and El-Hawary (1993) or Mak et al. (1997). So basically we apply a weighted least squares estimation of model (1). Then we use the corresponding residuals to estimate the volatility which is used to compute new weights for re-estimating model (1).

The considered model has a lot of regressors, and it is likely that some of them have no significant impact on the model, but rather increase spurious effects. The suggested lasso (least absolute shrinkage and selection operator) approach can handle this in an efficient way. It was introduced by Tibshirani (1996) in the context of shrinkage and selection in regression models and was recently applied by Hsu et al. (2008) and Ren and Zhang (2010) in the context of vector autoregressive models.

The weighted lasso estimator of the parameter vector $\boldsymbol{\theta}^i$ of Eqs. (1) or (7) is given by

$$\hat{\theta}_i = \underset{\theta\in\mathbb{R}^{p_i}}{\arg\min}\sum_{t=1}^{n}\left(Y_t^i - w_t^i\mathbf{X}_t^i\boldsymbol{\theta}^i\right)^2 + \lambda_{i,n}\sum_{j=1}^{p_i}|\theta_j| \tag{8}$$

with a weight vector $\mathbf{w}^i = (w_1^i, \ldots, w_n^i)$, p_i as the length of $\boldsymbol{\theta}^i$ and tuning parameter $\lambda_{i,n}$. In the initial step $\mathbf{w}^i$ is chosen to be 1. For the estimation we use the least angle regression (LARS) estimation algorithm of Efron et al. (2004).

As selection criterion we consider the Akaike information criterion (AIC) to estimate. This is asymptotically equivalent to the cross-validation in regression analysis as suggested in Efron et al. (2004)

In the next step we want to compute a new weight matrix $\mathbf{W} = (\mathbf{w}^1, \ldots, \mathbf{w}^d)$ based on the volatility estimates for $\boldsymbol{\sigma}_t^i$. Thus we have to estimate the volatility parameters from Eq. (4) using Eqs. (5) and (6).

Here we have to ensure that all parameters take non-negative values, as the regressors are all positive. To solve this non-negative least squares problem, we use the NNLS algorithm as described by Lawson and Hanson (1995) Note that this estimation technique can be seen as a parameter selection approach as well, as some parameters are likely estimated to be 0. This sparsity effect in a NNLS problem was recently analyzed by Meinshausen et al. (2013). Slawski et al. (2013) showed that the non-negative least squares approach is potentially superior to the positive lasso, a lasso approach with a positive parameter constraint.

The general estimation scheme is given by

1) Set the initial $d \times n$ dimensional weight matrix $\mathbf{W} = (1, \ldots, 1)$ and the iteration parameter $K = 1$.
2) Estimate Eq. (1) using LARS-lasso method with weights $\mathbf{W}$.
3) Estimate σ_t by Eqs. (5) and (6) with $|\hat{\varepsilon}_t|$ as absolute residuals from 2) using the NNLS algorithm.
4) Redefine $\mathbf{W} = (\mathbf{w}^1, \ldots, \mathbf{w}^d)$ by $\mathbf{w}^i = \left(\left(\hat{\sigma}_1^i\right)^{-2}, \ldots, \left(\hat{\sigma}_n^i\right)^{-2}\right)$ with $\hat{\boldsymbol{\sigma}}_t = \left(\hat{\sigma}_t^1, \ldots, \hat{\sigma}_t^d\right)$ as fitted values from 3).
5) Stop the algorithm, if a stopping criteria is satisfied, otherwise $K = K + 1$ and go back to 2).

Note that before performing the LARS algorithm on the weighted regressor matrix, we standardize their columns so that the columns of $\mathbf{X}^i$ from Eq. (7) have zero sample mean and sample variance 1, so that LASSO can act as a proper model selection algorithm. Additionally we

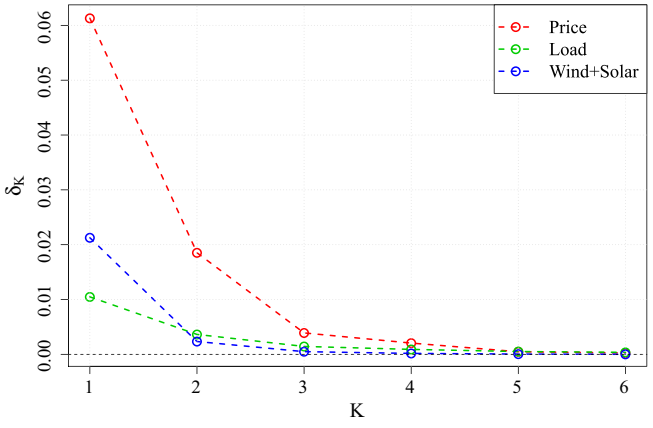

Fig. 6. δ_K^i for iteration $K \in \{1, \ldots, 6\}$ and $i \in \mathscr{I}$, estimated on the full available sample.

Table 5
Summary table of the amount of used parameters p_i of the model given in Eq. (1) and the TARCH model in Eq. (4) $i \in \mathscr{I}$.

Number of parameters	$\mu^i(t)$	$I_{i,\mathscr{P}}$	$I_{i,\mathscr{L}}$	$I_{i,\mathscr{R}}$	Seasonal $\phi^{i,j}$ coefficients	Sum (p_i)	α_i (TARCH)
$i = \mathscr{P}$	36	369	369	49	6×35	1033	402
$i = \mathscr{L}$	36	–	369	49	6×35	664	402
$i = \mathscr{R}$	37	–	–	361	5×36	578	397

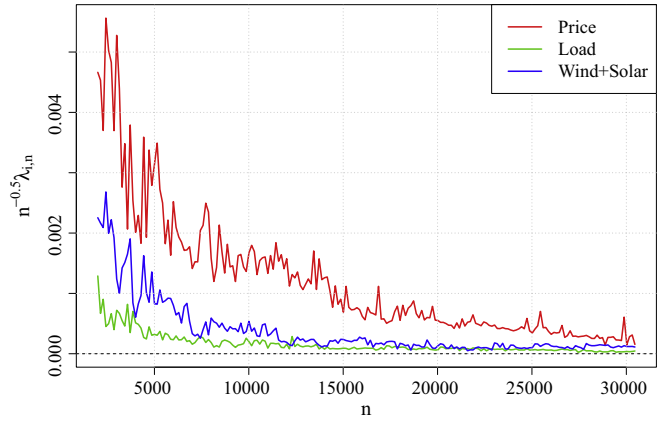

Fig. 7. $n^{-\frac{1}{2}}\lambda_{i,n}$ for a selection of various n.

could also replace $\hat{\boldsymbol{\sigma}}_t$ by $\hat{\tilde{\boldsymbol{\sigma}}}_t$ as they are the same up to the constant $\hat{\gamma}_i$. This does not change the result of the estimation algorithm, but we can improve the computation time slightly as we do not have to estimate $\boldsymbol{\gamma}^i$ by Eq. (6).

For the stopping criteria we suggest to look at the convergence of the $\boldsymbol{\sigma}^i$ within the algorithm. A reasonable criteria is to iterate until $\delta_K^i = \|\hat{\boldsymbol{\sigma}}_{K-1}^i - \hat{\boldsymbol{\sigma}}_K^i\| < \epsilon$ for all $i \in \mathcal{I}$. For $K = 1$ the $\boldsymbol{\sigma}_1^i$ is the homoscedastic estimate from the homoscedastic lasso estimation. For our application we choose δ_K^i as $n^{-1}\|\hat{\boldsymbol{\sigma}}_{K-1}^i - \hat{\boldsymbol{\sigma}}_K^i\|_1$ and $\epsilon = 0.001$. In the empirical study we noticed that this algorithm converges quickly (see Fig. 6), so a small number of iterations $K_{\max}$ seems to be sufficient for practitioners. In Fig. 6 we can observe that in the fifth iteration the change in $\hat{\boldsymbol{\sigma}}^i$ is less than ϵ. Hence, we decide to stop after $K_{\max} = 4$ iterations in the rest of our study.

The estimation technique used has the huge advantage of being fast, even if hundreds of potential regressors are included in the model. The computational complexity of the dominating LARS algorithm is the same as for a common OLS estimation, which is $\mathcal{O}(K_{\max} d p^2 n)$ for $p < n$ where $K_{\max}$ is the number of maximal iteration steps, d the dimension of the process, p the maximal number of regressors in Eq. (7), and n the number of observations. Here we see that the number of parameters p in the models has a quadratic impact, thus it is important to keep the parameter space manageable regarding the computation time. In our computations the iteratively reweighted lasso with $K_{\max} = 4$ takes only a few minutes on a 3 GHz computer, with n about 31,000, p about 3500, and $d = 3$.

The consistency and asymptotic behavior of $\hat{\boldsymbol{\theta}}_t$ and $\hat{\boldsymbol{\alpha}}_i$ for the process considered is not obvious. However, the lasso-type estimators in a heteroscedastic regression setting were recently analyzed in Nkurunziza and Lin (2012), Wagener and Dette (2012), and Wagener and Dette (2013).

There are also some asymptotic results for the lasso estimator in an autoregressive setting. Important results are given for univariate ARX processes by Wang et al. (2007) and for the VAR model by Hsu et al. (2008).

Even though there is a lack of theoretical justification for the asymptotic properties of the used estimators, we will assume asymptotic normality for $\hat{\theta}_i$ and $\hat{\boldsymbol{\alpha}}_i$. For $\hat{\theta}_i$ we assume the asymptotic behavior as in Wagener and Dette (2012). Following their result we have $\sqrt{n}\left(\hat{\boldsymbol{\theta}}_i(1) - \boldsymbol{\theta}_i^*(1)\right) \to N(\boldsymbol{\zeta}_i(\boldsymbol{\theta}_i^*), \boldsymbol{\Gamma}_i(\boldsymbol{\theta}_i^*))$ in distribution where $\boldsymbol{\theta}_i^*$ is the true parameter vector, $\boldsymbol{\theta}_i(1)$ and $\boldsymbol{\theta}_i^*(1)$ the corresponding non-zero parts, $\boldsymbol{\zeta}_i(\boldsymbol{\theta}_i^*)$ as asymptotic mean vector, and $\boldsymbol{\Gamma}_i(\boldsymbol{\theta}_i^*)$ the asymptotic covariance matrix, if $n^{-\frac{1}{2}}\lambda_{i,n} \to \lambda_{i,0} \geq 0$. For the mean and the covariance matrix we have $\boldsymbol{\zeta}_i(\boldsymbol{\theta}_i^*) = -\frac{\lambda_{i,0}}{2}\mathbf{G}_i^{-1}\operatorname{sign}(\boldsymbol{\theta}_i^*(1))$ and $\boldsymbol{\Gamma}_i(\boldsymbol{\theta}_i^*) = \boldsymbol{G}_i^{-1}\boldsymbol{G}_i^{\boldsymbol{W}}\boldsymbol{G}_i^{-1}$ with $\boldsymbol{G}_i$ as limit of the Gramian $\boldsymbol{G}_{n,i} = n^{-1}\boldsymbol{X}_i^\top \boldsymbol{X}_i$, $\boldsymbol{G}_i^{\boldsymbol{W}}$ as limit of the weighted Gramian $\boldsymbol{G}_{n,i}^{\boldsymbol{W}} = n^{-1}\boldsymbol{X}_i^\top \boldsymbol{W}_i \boldsymbol{X}_i$, weight matrix $\mathbf{W}_i = \operatorname{diag}((\sigma_1^i)^{-2}, \ldots, (\sigma_n^i)^{-2})$, and sign as signum function.

We see that the asymptotic behavior depends linear on the unknown limit $\lambda_{i,0}$. For analyzing the asymptotic in our situation we estimated the model for many different n ranging from 2000 to our maximum sample size and computed the corresponding $\lambda_{n,i}$. In Fig. 7 we plotted n against $n^{-\frac{1}{2}}\lambda_{i,n}$.

There we can observe that $n^{-\frac{1}{2}}\lambda_{i,n}$ goes clearly to zero as n increases. Thus the assumption $\lambda_0 = 0$ seems to be appropriate, and the zero vector $\mathbf{0}$ is a reasonable choice as an estimator for the mean vector $\boldsymbol{\zeta}_i(\boldsymbol{\theta}_i^*)$. In addition the covariance matrix can be estimated by the common plug-in estimator $\hat{\boldsymbol{\Gamma}}_i = n(\boldsymbol{X}_i^\top \boldsymbol{X}_i)^{-1}\boldsymbol{X}_i^\top \hat{\boldsymbol{W}}_i \boldsymbol{X}_i (\boldsymbol{X}_i^\top \boldsymbol{X}_i)^{-1}$.

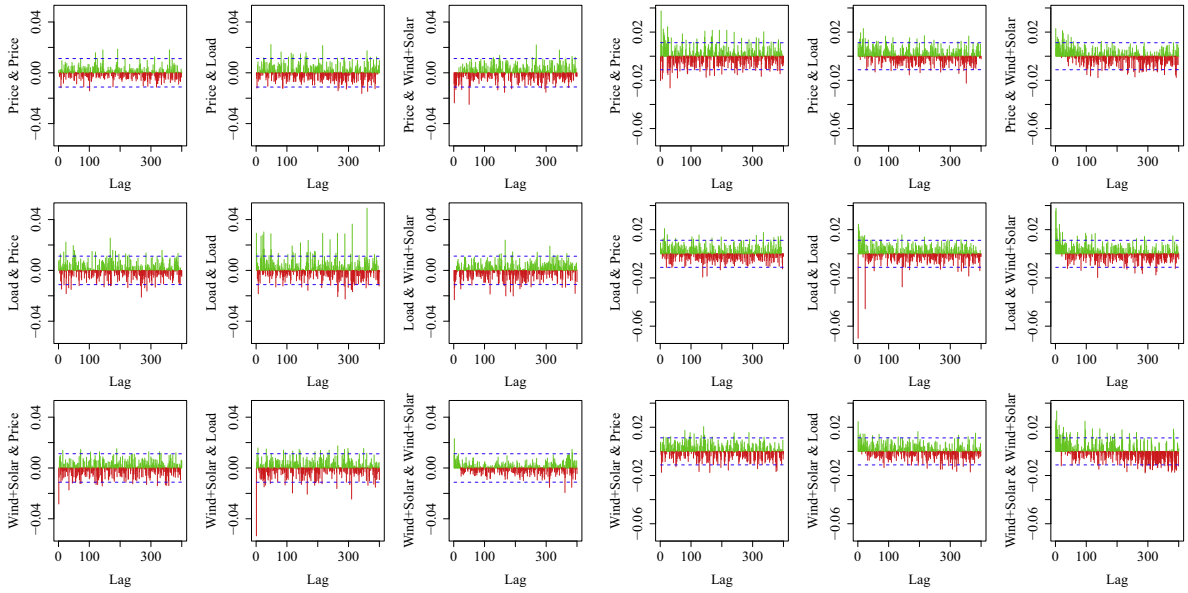

Fig. 8. Sample autocorrelation function of $_t$ (a) and $|_t|$ (b).

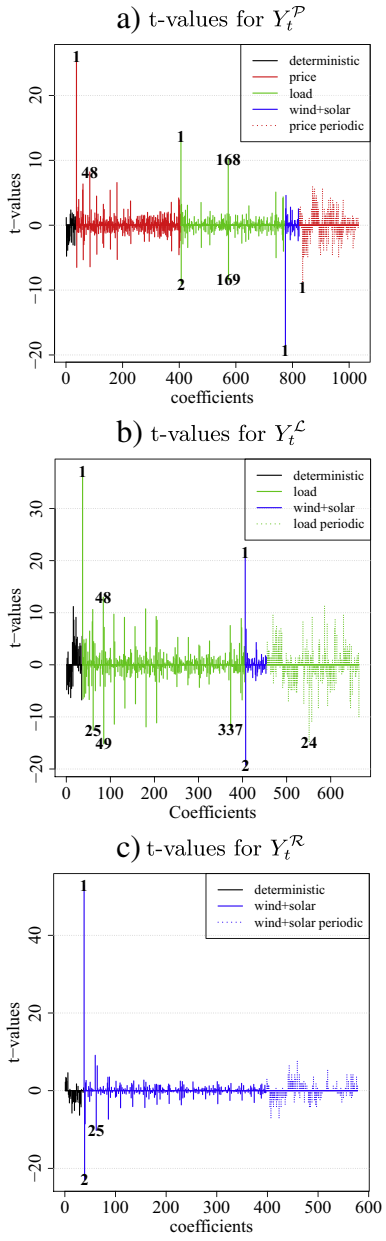

Fig. 9. t-values of the estimated model for $\mathbf{Y}_t$ distinguished by deterministic, price, load, wind + solar and periodic coefficients.

5. Results

We performed the mentioned estimation technique on the full data set. For the model diagnostic the sample autocorrelation function (ACF) of the estimated standardized residuals and their absolute values are given in Fig. 8.

The left 3×3 matrix depicts how the standardized residuals of one time series correlate with the lagged standardized residuals of the same or any other of the three time series. The right 3×3 matrix shows the same relation for the absolute standardized residuals. For instance, the left upper ACF shows the correlation of the standardized residuals of the price with its own lagged standardized residuals. The illustration right next to it, Price & Load, shows the correlation of the standardized residuals of the price with the lagged standardized residuals of the load. Hence, we can see that the independence assumption for $\mathbf{Z}_t$ seems to be satisfied for most of the relations. However, the strongest serial correlation structure can be observed for the standardized residuals $Z_t^{\mathscr{L}}$ of the load. There could be several reasons for the remaining correlation structure. For example, we may simply did not choose enough parameters to cover the complex dependence structure, so an enlargement of $I_{\mathscr{L},\mathscr{L}}$ or $L_{\mathscr{L},\mathscr{L}}$ or a reduction of $d_{\mathcal{K}}$ might help. It could be also possible that there might be non-linear or interactions effects that were not considered so far.

Further, the sample autocorrelation of $|\mathbf{Z}_t|$ is mainly zero, so that $\mathbf{Z}_t$ seems to be quite homoscedastic. Thus the TARCH approach seems to be appropriate to explain the conditional heteroscedasticity. Nevertheless, there is still space for improvements.

Given the estimated model (1) we can evaluate the t-values of every coefficient given by their estimate over its standard error which can be estimated under the asymptotic normality assumption. Roughly speaking we can say, that the larger the absolute t-value of a coefficient, the more significant is its impact on the process. The evaluated t-values are given in Fig. 9, distinguished into the groups introduced at the end of Section 3.

It is worth mentioning that the load and wind and solar feed-in seem to have a strong impact on the actual electricity price, $Y_t^{\mathcal{P}}$. The most recent past load values (lags 1 and 2) seem to be very important as well as the values one week ago (lags 168, 169). In contrast to this, wind and solar power has a vast impact only in the short-run at lag 1. For the load dependency structure, $Y_t^{\mathscr{L}}$, we receive that the first autoregressive lag and the subsequent daily lags tend to be of great importance. Interestingly, lag 337, which accounts for effects two weeks ago, seem to also have a special importance. But there is also a strong relationship concerning the renewable energy feed-in, which is especially high for lags 1 and 2. For the solar and wind power feed-in, $Y_t^{\mathscr{R}}$, it is interesting that the first two autoregressive lags clearly dominate all the other larger lags. However, the daily lagged variables still play an important role.

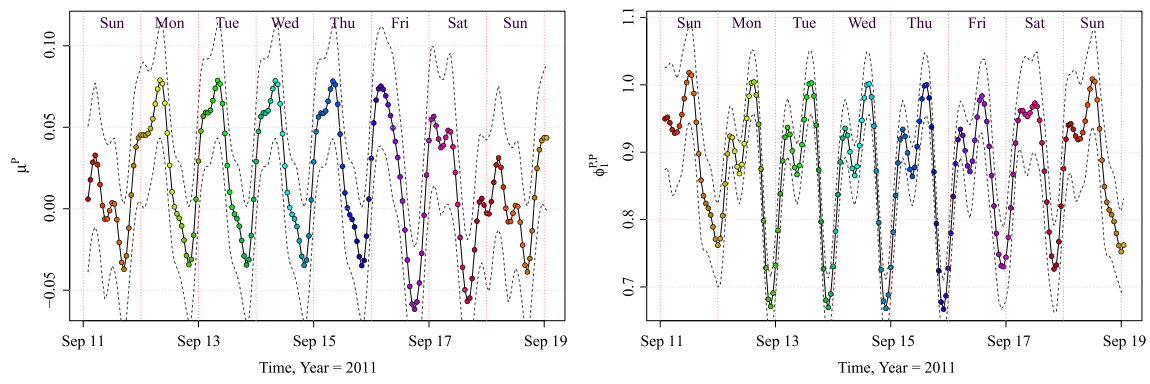

Fig. 10. $\hat{\mu}^{\mathcal{P}}(t)$ and $\hat{\phi}_1^{\mathcal{P},\mathcal{P}}(t)$ within 8 days with approximate 95% confidence intervals.

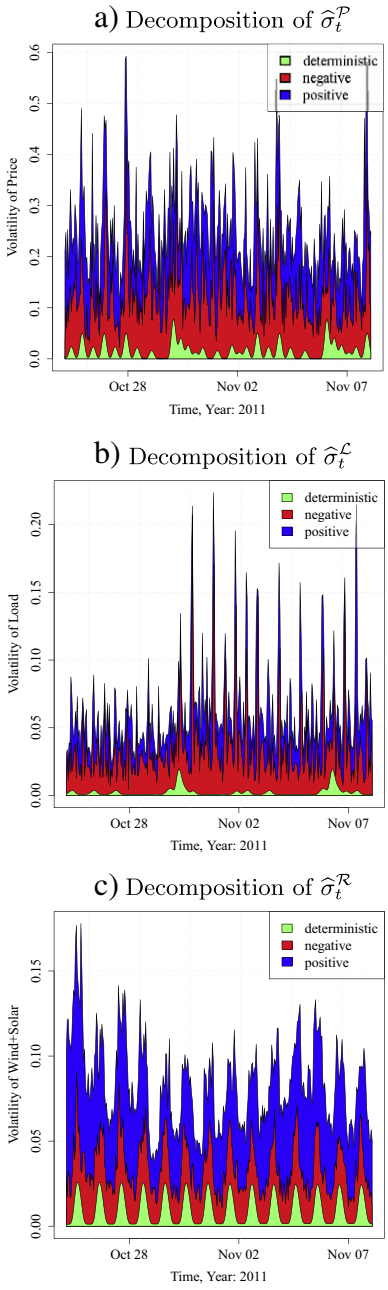

Fig. 11. Decomposition of the estimated volatilities $\hat{\sigma}_t^i$ into their deterministic seasonal part, the part influenced by positive residuals and positive ones for the price, load and wind + solar.

In general we can observe that the dependency structure is quite complex. Therefore we conclude that the chosen set of lags $I_{i,j}$ and $L_{i,i}$, though having thousands of parameters, seem to be reasonable. This is also backed by the results of the parameter selection algorithm. For the price process 82.2 % of the chosen parameters were included. The load process was set to 92.8 % and wind and solar to 76.6 % of the maximum amount of parameters. Especially the high percentage for the load indicates that an enlargement of the parameter set is likely to improve the results.

Assuming asymptotic normality allows us to compute several confidence intervals and bands. For example, for $\boldsymbol{\mu}(t)$ and time varying autoregressive parameters $\phi_k^{i,j}(t)$. We illustrate the behavior of $\mu^{\mathcal{P}}(t)$ and $\phi_1^{\mathcal{P},\mathcal{P}}(t)$ in Fig. 10. From this illustration it can be seen that all coefficients have distinct weekly patterns and a seasonal structure.

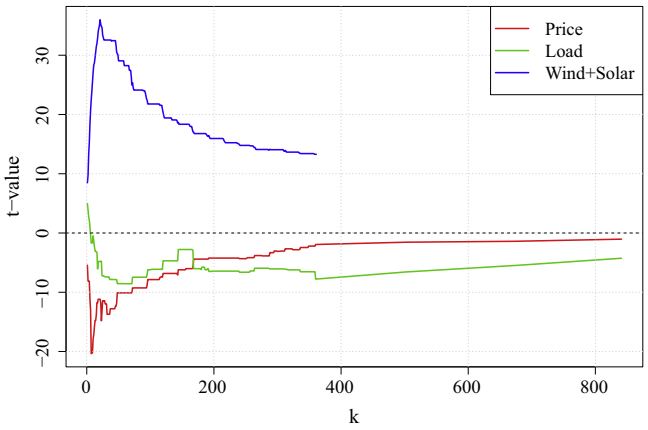

Fig. 12. t-Values of $A_{i,k}$ for $i \in \mathcal{I}$.

Moreover, analyzing the volatility structure determined by $\boldsymbol{\sigma}_t$ yields promising results. Using the TARCH model for ε_t allows us to decompose $\boldsymbol{\sigma}_t$ into three components. The first one is determined by the periodic coefficient $\boldsymbol{\alpha}_0(t)$. It describes the deterministic part of the volatility and is automatically a lower bound for $\boldsymbol{\sigma}_t$. The second and third components describe the residual impact of the positive and negative residuals, respectively. Note that in an ARCH model, both components are equal. The decomposition of the estimated volatilities $\hat{\sigma}_t^i$ is given in Fig. 11. For illustrative purposes, we selected approximately two weeks of October/November 2011. It can be seen that the volatility of each of the three time series is influenced by a deterministic seasonal component. Moreover, for the price process, we cannot observe a strong leverage effect, the impact of past negative residuals is almost equal to the impact of the past positive ones. This is represented by the area marked as *negative* in the left picture being approximately equal to the area marked as *positive*. This relation is obviously different for load time series. Here we observe a strong leverage effect. The impact of negative residuals is higher than the impact of positive residuals. For the wind and solar feed-in an inverse leverage effect is present, as the *positive* area clearly dominates the *negative* area.

Furthermore, we can perform tests to check the significance of the leverage effect. This will be especially interesting for the price series, as judging from our subset of October/November 2011 it was not clear whether a significant effect occurred. Therefore we evaluate several tests $H_0 : A_{i,k} = \sum_{j=1}^{k} \alpha_j^{+,i} - \alpha_j^{-,i} = 0$ for $i \in \mathcal{I}$. The k indicates the lags up to which the leverage effect is taken into account.

Fig. 12 provides the t-values of the corresponding test for all possible k. In order to construct the different values for k we simply cut the remaining model parameters of the TARCH-part after k and performed the test described above. We therefore explicitly state that we did not re-estimate the model for every k.

It is obvious that the wind and solar energy has clearly positive t-values with its peak at about a day. Those positive t-values indicate a strong inverse leverage effect. For the load time series the leverage effect seems to point in the opposite direction. Including all model parameters we can state, that there is a significant standard leverage effect within the data. This is also supported by the t-value for the load in Table 6. However, the picture becomes fuzzy when the electricity price is considered. Judging by the figure it can be obtained, that the electricity prices firstly experiences a positive leverage effect, which

Table 6
Test for the leverage effect $H_0 : A_{i,k} = 0$, including all estimated parameters.

	$\hat{A}_{i,k=\max}$	$\hat{\sigma}\left(\hat{A}_{i,k=max}\right)$	t-Value	p-Value
Price ($i = \mathcal{P}$)	−0.0561	0.0540	−1.0394	0.2986
Load ($i = \mathcal{L}$)	−0.1780	0.0417	−4.2666	0.0000
Wind + Solar ($i = \mathcal{R}$)	0.8427	0.0635	13.2689	0.0000

Table 7
Tests for effects of the regressors on the price, two-tailed test.

load on price				Wind + solar on price			
$\hat{\Phi}_{\mathcal{P},\mathcal{L}}$	$\hat{\sigma}\left(\hat{\Phi}_{\mathcal{P},\mathcal{L}}\right)$	t-Value	p-Value	$\hat{\Phi}_{\mathcal{P},\mathcal{R}}$	$\hat{\sigma}\left(\hat{\Phi}_{\mathcal{P},\mathcal{R}}\right)$	t-Value	p-Value
−0.064	0.180	−0.356	0.721	−0.732	0.082	−8.921	0.000

diminishes after a certain amount of time. When all parameters up to k are considered, the leverage effect even becomes insignificant, as can be seen in Table 6. However, applying a variance model which can account for a leverage effect is still important also for electricity prices, as only those models can account for such a complex structure with different signed impacts of different lags on the volatility.

In the same way we can perform an approximate test as to whether an increasing load or wind-solar feed-in leads to a significantly increasing or decreasing price in the long run. These tests have the null hypothesis $H_0 : \Phi_{\mathcal{P},\mathcal{L}} = \sum_{k\in I_{\mathcal{P},\mathcal{L}}} \phi_k^{\mathcal{P},\mathcal{L}} = 0$ and $H_0 : \Phi_{\mathcal{P},\mathcal{R}} = \sum_{k\in I_{\mathcal{P},\mathcal{R}}} \phi_k^{\mathcal{P},\mathcal{R}} = 0$. Applying this test to the data (see Table 7) it can be concluded that there is only a small long run impact of the load to the electricity price, which is statistically not significant. In contrast, an increasing in the amount of wind or solar power feed-in causes a decrease in the price. Taking the standardization of Y_t^i by its mean and standard deviation before the estimation of Eq. (1) into account we can quantify these results. An increase of 1 GWh in the load changes the price by $-0.108(\pm 0.499)_{\frac{\text{EUR}}{\text{MWh}}}$[5] and an increase of 1 GWh wind or solar power feed-in changes the price by $-2.031(\pm 0.375)_{\frac{\text{EUR}}{\text{MWh}}}$.

This result concerning the wind and solar feed-in is consistent with the recent literature on quantifying the impact of renewable energy in Germany, see e.g. Würzburg et al. (2013).

Moreover, we estimated a linear trend in $Y_t^{\mathcal{P}}$ which is negative and significantly different from zero, so over time, the price seems to decrease. But this effect is not that strong: over one year, the price changes by $-0.0200(\pm 0.0069)_{\frac{\text{EUR}}{\text{MWh}}}$. Nevertheless, we have to be careful with extrapolating a linear trend far in the future, as it is economically impossible for such a negative trend to last forever.

6. Forecasting

Given the estimated model (1) of the sample $(\mathbf{Y}_1, \ldots, \mathbf{Y}_n)$ we can easily carry out a forecast. Let $\mathbf{Y}_t = g(\mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \ldots)$ be the representation of Eq. (1). Then we can compute $\hat{\mathbf{Y}}_{n+h}$ iteratively by

$$\hat{\mathbf{Y}}_{n+h} = g\left(\hat{\mathbf{Y}}_{n-1+h}, \hat{\mathbf{Y}}_{n-2+h}, \ldots\right)$$

and defining $\hat{\mathbf{Y}} := \mathbf{Y}_t$ for $t \le n$.

We performed a forecasting study, where we choose subsequences of the $n = 31465$ observations. So we choose a data part H_l that is approximately two years long (exactly 18481 observations = 110 weeks + 1 h), starting at observation l. The l is chosen so that the observed sample ends with an observation for the electricity price from 23 h to 24 h. Finally, we performed a complete estimation on the data sample H_l and estimate the next $h = 672$ h, that is four weeks. This method provides a fair estimation technique, as we use the same amount of past observations for every forecast. Given our sample the procedure provides $N = 506$ observations, while $\hat{\mathbf{Y}}_{h,k}$ and $\mathbf{Y}_{h,k}$ for $k \in \{1, \ldots, N\}$ denote the corresponding predicted values and observation, respectively. Note that we are forecasting the three-dimensional process $\mathbf{Y}_t$. However, in this paper we are interested in the price process $Y_t^{\mathcal{P}}$, therefore the further discussion will focus on this process.

The first benchmark we consider is the homoscedastic solution of model (1). We get its forecasts automatically by taking the estimated parameters after the first iteration step within the iterative estimation algorithm. Thus, we can directly see the impact of the considered TARCH part.

As other own benchmarks we consider the weekly persistent process $Y_t^{\mathcal{P}} = Y_{t-168}^{\mathcal{P}}$ as persistent model, such as two AIC selected (V)AR processes with time varying mean. Their model is given by

$$\boldsymbol{X}_t = \boldsymbol{\mu}_j 1_{\{t\in(168\mathbb{N}+j)\}} + \sum_{k=1}^{p}\left(\boldsymbol{\Phi}_k \boldsymbol{X}_{t-k} - \boldsymbol{\mu}_j 1_{\{t\in(168\mathbb{N}+j)\}}\right) + \varepsilon_t$$

where $j \in \{1, \ldots, 168\}$. We consider the univariate choice $\mathbf{X}_t = Y_t^{\mathcal{P}}$ and the two-dimensional incorporating the load as well, so $\mathbf{X}_t = \left(Y_t^{\mathcal{P}}, Y_t^{\mathcal{L}}\right)$. We also tried to include the wind and solar feed-in but the out-of-sample performance got worse. We estimated the models in a two step approach, first removing the weekly mean and second estimating a (V)AR process via Gaussian AIC selection, where the 168 parameters for the mean are ignored. For the estimation process we solve the Yule–Walker equations that provide a guaranteed stationary solution. As maximal possible order for the univariate process, we choose $p_{\max} = 1210$ and for the two-dimensional one $p_{\max} = 555$. The estimation of the processes is very fast and done in a few seconds. However, in our empirical results, the order p of the AR process is usually automatically chosen to be about 800, whereas the in the two-dimensional VAR case, it is usually about 400, so both cover the weekly mean and conditional mean behavior.

Moreover, we tried to use as many models from the recent literature for a benchmark as possible. But unfortunately finding the right competitor is difficult for several reasons. First of all many authors only consider positive observations or delete some chunks of the data (e.g. outliers, holidays, etc.). Second, some of them consider information from random regressors such as the load as known, which in a real world situation would not be the case. And finally there are some models, especially from machine learning, that simply require an enormous computational time. Furthermore, note that feed-forward neural networks with one hidden layer and lagged process $\mathbf{Y}_t$ as input acts very similar to an AR(p) process if the chosen lags in the input layer are covered in an AR(p) process. But, as mentioned, neural networks are extremely time consuming, due to the learning phase, which is often based on random selections.

However, we consider three benchmarks from the literature that have been applied to electricity prices. First, an ARMA(5,1) model with trend as well as annual, weekly and daily cycles as suggested as one of the best models in Keles et al. (2012). Note that their daily cycles vary over the seasons of the year. Moreover, we consider the functional data analysis approach from Liebl (2013). But he also modifies the data: he removes outliers and excludes holidays and weekends. In fact, he models the prices by estimating the merit order curve under consideration of the load subtracted by the wind power feed-in. But for our data we noticed that his model gains better results when only the load is used. Furthermore, in his studies, two functional principal components were sufficient to model the data well. But we got better results by using three principal components. In the original paper the loading coefficients were predicted using a SARIMA model with periodicity five, whereas we use the same model with a periodicity seven as we are not ignoring the weekends. Moreover, we remark that the usage of the basic model of Liebl (2013) as a benchmark to our model is limited, as his approach was tailor-made for the exclusion of special days like weekends and heavy outliers. It may therefore be the case that our reported model performance for his model is only due to the violation of some of his basic assumptions. As a third benchmark from the existing literature we employ the wavelet-ARIMA approach from Conejo et al. (2005). Such a model is often used as a benchmark in electricity spot price forecasting. In our application we use the Daubechies 4 wavelet. For modeling the coefficients of the wavelet decomposition we choose ARIMA(12,1,1) processes to capture their autoregressive structure.

[5] The term in parentheses gives the symmetric 90% confidence interval.

Table 8
$MAE_h^{\mathcal{P}}$ and $MMAE_h^{\mathcal{P}}$ in EUR/MWh of several models: pVAR = homoscedastic lasso solution of model (1), pVAR-TARCH = proposed model. Best model is highlighted with bold font.

	pVAR-TARCH		pVAR		μ-AR(p)		μ-VAR(p)		Persistent		Keles et al.		Liebl		Conejo et al.	
h	$MAE_h^{\mathcal{P}}$	$MMAE_h^{\mathcal{P}}$	$MAE_h^{\mathcal{P}}$	$MMAE_h^{\mathcal{P}}$	$MAE_h^{\mathcal{P}}$	$MMAE_h^{\mathcal{P}}$	$MAE_h^{\mathcal{P}}$	$MMAE_h^{\mathcal{P}}$	$MAE_h^{\mathcal{P}}$	$MMAE_h^{\mathcal{P}}$	$MAE_h^{\mathcal{P}}$	$MMAE_h^{\mathcal{P}}$	$MAE_h^{\mathcal{P}}$	$MMAE_h^{\mathcal{P}}$	$MAE_h^{\mathcal{P}}$	$MMAE_h^{\mathcal{P}}$
1	**3.20**	**3.20**	3.58	3.58	3.29	3.29	3.25	3.25	8.40	8.40	3.46	3.46	6.24	6.24	5.69	5.69
4	**4.76**	**4.09**	5.52	4.64	5.38	4.52	5.40	4.52	10.91	9.75	6.79	5.25	7.28	6.83	6.76	6.35
8	**5.57**	**4.60**	6.00	5.31	6.30	5.01	6.32	5.04	10.68	9.99	11.27	6.66	11.96	7.81	12.39	7.48
12	**7.02**	**5.20**	7.17	5.73	7.43	5.67	7.50	5.71	11.48	10.37	9.55	7.90	16.52	10.42	10.21	8.78
16	**6.79**	**5.64**	7.15	6.09	7.13	6.09	7.11	6.13	10.62	10.55	9.71	8.34	15.31	11.87	10.83	9.21
20	7.53	**5.97**	**7.48**	6.37	8.01	6.43	7.98	6.44	10.29	10.60	9.37	8.68	12.50	12.13	10.66	9.58
24	5.58	**5.95**	5.90	6.31	**5.53**	6.37	5.65	6.38	7.97	10.21	6.58	8.42	7.55	11.58	7.17	9.30
168	**6.56**	**7.89**	7.00	8.29	6.74	8.04	6.97	8.03	7.92	10.18	7.07	9.51	7.83	13.69	8.79	11.15
672	7.19	**8.76**	7.31	9.01	8.83	9.56	8.95	9.59	8.89	11.66	**7.00**	9.77	8.82	14.99	11.22	13.33

As a performance measure for our forecasting study, we do not use MAPE or WMAPE as is often done in the literature, because the $MAPE_h^{\mathcal{P}} = \frac{1}{N}\sum_{k=1}^{N}\left|\frac{Y_{h,k}^{\mathcal{P}} - \hat{Y}_{h,k}^{\mathcal{P}}}{Y_{h,k}^{\mathcal{P}}}\right|$ is obviously pointless for data that can take zero values. Instead, we consider the mean absolute forecast error for the prediction time h ($\mathbf{MAE}_h$) as well as the mean of the mean absolute forecast error up to a forecast horizon of h ($\mathbf{MMAE}_h$). They are given by

$$\mathbf{MAE}_h = \frac{1}{N}\sum_{k=1}^{N}|\mathbf{Y}_{h,k} - \hat{\mathbf{Y}}_{h,k}| \quad \text{and} \quad \mathbf{MMAE}_h = \frac{1}{hN}\sum_{j=1}^{h}\sum_{k=1}^{N}|\mathbf{Y}_{j,k} - \hat{\mathbf{Y}}_{j,k}|.$$

As main criterion to compare different models we suggest the $\mathbf{MMAE}_{24}$, as it estimates the expected average absolute error for the whole next trading day, which is especially for the practical application of relevance. The computed $MAE_h^{\mathcal{P}}$ and $MMAE_h^{\mathcal{P}}$ for the considered models with selected prediction horizons are given in Table 8. The evolution of $MAE_h^{\mathcal{P}}$ and $MMAE_h^{\mathcal{P}}$ with increasing h is presented in Fig. 13.

It is remarkable that our proposed model can outperform every other one, especially these that are used in the literature. Remember that we compare pure out-of-sample methods, that use no information from the prediction time during the estimation. Moreover, we noticed that many models are not able to outperform simple AIC selected AR(p) or VAR(p) models, so e.g. Keles et al. (2012) and Liebl (2013). Such benchmarks have not been considered in most of the literature so far, even though they are simple and fast to estimate. The relatively good performance is likely due to the highly carefully chosen model order, so that hundreds of variables are included in the model, which can cover well the behavior. This amount of chosen parameters is in the same range as for our reweighted lasso procedure.

Furthermore, we use Monte Carlo methods to compute the prediction bands conditioned on the given data. Thus, we can perform a residual based bootstrap on the standardized residuals $_t$ to simulate $\mathbf{Y}_t$. So we resample from $_t$ to simulate Σ_t and afterwards $\mathbf{Y}_t$ to compute the prediction intervals.

For illustrative purposes, we performed a prediction for the sample that is given in Fig. 1, which includes a public holiday. So the considered estimation period is about two years and the prediction horizon is 192 h. The corresponding point estimates, such as the estimated conditional mean and the 90 % and 99 % prediction intervals are given in Fig. 14. The prediction bands are computed by evaluating the symmetric conditional value at risk (VaR) resp. quantile levels.

Interestingly, the forecasting performs quite well. For example, Wednesday the 3rd October is an official holiday, so the forecasted values are significantly smaller than those, for example, for the 4th October, even though in this case the forecast still overestimates the real value.

7. Summary and conclusion

The paper presented a model for the hourly electricity price of the European Power Exchange for Germany and Austria. A periodic VAR-TARCH approach was proposed in order to capture the specific price movements. This model is able to deal with the most difficult challenges in modeling electricity prices. Namely, it considers: the mean-reversion of prices, the seasonality of the data, the possibility of vast price spikes, negative prices, periodicity, time dependent variance, the leverage effect, the impact of renewable energy, the impact of the electricity load, calendar effects like holidays, and time shifts due to daylight saving time.

For the necessary simultaneous modeling of the price, load, wind and solar power, we used a modern estimation approach, which is efficient and turned out to enable rapid estimations. By fitting our model to the before mentioned data sets we were able to show insides towards the leverage effect of our considered time series. Moreover, we provided

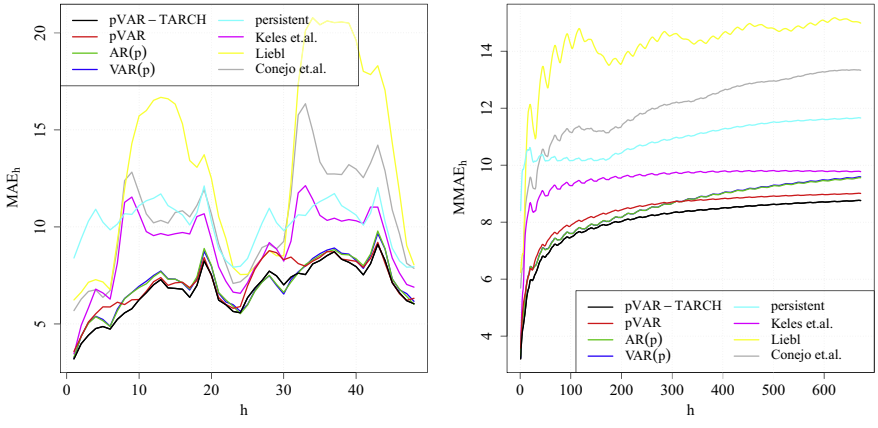

Fig. 13. $MAE_h^{\mathcal{P}}$ and $MMAE_h^{\mathcal{P}}$ in EUR/MWh of several models and different forecasting periods.

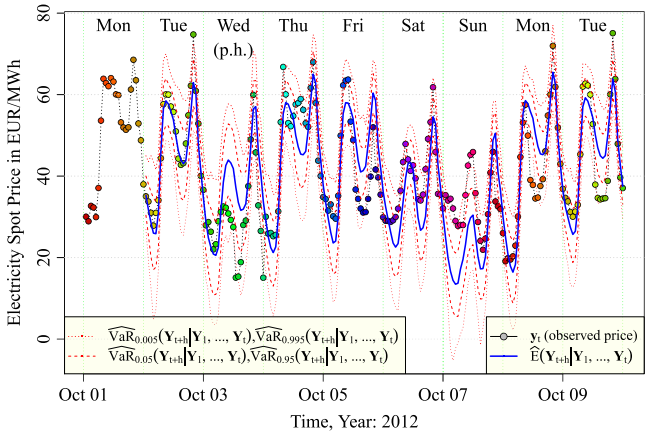

Fig. 14. Electricity price prediction with 90% and 99% prediction bands for $h \in \{1, \ldots, 192\}$ using a Monte-Carlo sample size of 1000. The 3rd of October is a public holiday (German unity).

evidence for the effect of renewable energy in reducing the price of electricity. Our study showed that an increase in combined wind and solar power of 1 GWh leads to a decrease in price of 2.03 $\frac{\text{EUR}}{\text{MWh}}$.

An extensive forecasting study showed that in MAE and MMAE our model outperforms every model which was used as a benchmark, including recently published models in the literature. For the purpose of forecasting, no future knowledge was necessary: every forecast was made exclusively with out-of-sample data. Due to its high efficiency and fast computing time, our model may be of interest for researchers who want to use it as a benchmark for their own model as well as for energy companies who need to forecast electricity prices.

Nevertheless, there is still much research left for the future. For instance, as the markets of some European countries are strongly integrated, measuring the import and export feeds may lead to better estimation results and provide guidance for political decisions.

It is also debatable whether our model can be reasonably applied to other electricity markets. We notice that the characteristics of the energy portfolio and working day structure of Germany had significant impacts on the price process. But those characteristics are not necessarily possessed by other markets. Therefore, introducing a more comprehensive model may be an appropriate future task.

Another direction for future work could also be a more detailed examination of the German energy portfolio, by considering the time series of prices for other energy resources, e.g. coal or gas. Detecting and using the cointegration between those time series may also lead to an improvement in the estimation results.

From the statistical point of view the considered model can be extended as well. So it is possible to include non-linear effects and interactions or change points within the variables. Further one might allow more parameter to vary periodically, as, for instance, it might be that the TARCH parameters change over time.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.eneco.2014.10.012.

References

Aggarwal, S.K., Saini, L.M., Kumar, A., 2009. Electricity price forecasting in deregulated markets: a review and evaluation. Int. J. Electr. Power Energy Syst. 31 (1), 13–22.

Black, F., 1976. Studies in stock price volatility changes. Proceedings of the Business and Economics Statistics Section. American Statistical Association, pp. 177–181.

Bordignon, S., Bunn, D.W., Lisi, F., Nan, F., 2013. Combining day-ahead forecasts for British electricity prices. Energy Econ. 35, 88–103.

Bosco, B.P., Parisio, L.P., Pelagatti, M.M., 2007. Deregulated wholesale electricity prices in Italy: an empirical analysis. Int. Adv. Econ. Res. 13 (4), 415–432.

Bowden, N., Payne, J.E., 2008. Short term forecasting of electricity prices for MISO hubs: evidence from ARIMA–EGARCH models. Energy Econ. 30 (6), 3186–3197.

Cancelo, J.R., Espasa, A., Grafe, R., 2008. Forecasting the electricity load from one day to one week ahead for the Spanish system operator. Int. J. Forecast. 24 (4), 588–602.

Christensen, T., Hurn, A., Lindsay, K., 2012. Forecasting spikes in electricity prices. Int. J. Forecast. 28 (2), 400–411.

Cifter, A., 2013. Forecasting electricity price volatility with the Markov-switching GARCH model: evidence from the Nordic electric power market. Electr. Power Syst. Res. 102, 61–67.

Conejo, A.J., Plazas, M.A., Espinola, R., Molina, A.B., 2005. Day-ahead electricity price forecasting using the wavelet transform and ARIMA models. IEEE Trans. Power Syst. 20 (2), 1035–1042.

Cruz, A., Muñoz, A., Zamora, J.L., Espínola, R., 2011. The effect of wind generation and weekday on Spanish electricity spot price forecasting. Electr. Power Syst. Res. 81 (10), 1924–1935.

Edenhofer, O., Hirth, L., Knopf, B., Pahle, M., Schlömer, S., Schmid, E., Ueckerdt, F., 2013. On the economics of renewable energy sources. Energy Econ. 40, 12–23.

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., 2004. Least angle regression. Ann. Stat. 32 (2), 407–499.

Escribano, A., Ignacio Peña, J., Villaplana, P., 2011. Modelling electricity prices: international evidence. Oxf. Bull. Econ. Stat. 73 (5), 622–650.

Eydeland, A., Wolyniec, K., 2003. Energy and Power Risk Management: New Developments in Modeling, Pricing, and Hedging. John Wiley & Sons, NJ, USA.

Fanone, E., Gamba, A., Prokopczuk, M., 2013. The case of negative day-ahead electricity prices. Energy Econ. 35, 22–34.

Guthrie, G., Videbeck, S., 2007. Electricity spot price dynamics: beyond financial models. Energy Policy 35 (11), 5614–5621.

Harvey, A., Koopman, S.J., 1993. Forecasting hourly electricity demand using time-varying splines. J. Am. Stat. Assoc. 88 (424), 1228–1236.

Hsu, N.-J., Hung, H.-L., Chang, Y.-M., 2008. Subset selection for vector autoregressive processes using lasso. Comput. Stat. Data Anal. 52 (7), 3645–3657.

Huisman, R., Kilic, M., 2012. Electricity futures prices: indirect storability, expectations, and risk premiums. Energy Econ. 34 (4), 892–898.

Huisman, R., Stradnic, V., Westgaard, S., 2013. Renewable energy and electricity prices: indirect empirical evidence from hydro power. IEB Working Paper N. 2013/024. Available at SSRN.

Karakatsani, N.V., Bunn, D.W., 2008. Forecasting electricity prices: the impact of fundamentals and time-varying coefficients. Int. J. Forecast. 24 (4), 764–785.

Keles, D., Genoese, M., Möst, D., Fichtner, W., 2012. Comparison of extended mean-reversion and time series models for electricity spot price simulation considering negative prices. Energy Econ. 34 (4), 1012–1032.

Keles, D., Genoese, M., Möst, D., Ortlieb, S., Fichtner, W., 2013. A combined modeling approach for wind power feed-in and electricity spot prices. Energy Policy 59, 213–225.

Knittel, C.R., Roberts, M.R., 2005. An empirical examination of restructured electricity prices. Energy Econ. 27 (5), 791–817.

Koopman, S.J., Ooms, M., Carnero, M.A., 2007. Periodic seasonal Reg-ARFIMA–GARCH models for daily electricity spot prices. J. Am. Stat. Assoc. 102 (477), 16–27.

Kristiansen, T., 2012. Forecasting Nord Pool day-ahead prices with an autoregressive model. Energy Policy 49, 328–332.

Lawson, C.L., Hanson, R.J., 1995. Solving Least Squares Problems. SIAM.

Liebl, D., 2013. Modeling and forecasting electricity spot prices: a functional data perspective. Ann. Appl. Stat. 7 (3), 1562–1592.

Liu, H., Shi, J., 2013. Applying ARMA–GARCH approaches to forecasting short-term electricity prices. Energy Econ. 37, 152–166.

Mak, T., Wong, H., Li, W., 1997. Estimation of nonlinear time series with conditional heteroscedastic variances by iteratively weighted least squares. Comput. Stat. Data Anal. 24 (2), 169–178.

Mbamalu, G., El-Hawary, M., 1993. Load forecasting via suboptimal seasonal autoregressive models and iteratively reweighted least squares estimation. IEEE Trans. Power Syst. 8 (1), 343–348.

Meinshausen, N., et al., 2013. Sign-constrained least squares estimation for high-dimensional regression. Electron. J. Stat. 7, 1607–1631.

Nkurunziza, S., Lin, E.Y.Y., 2012. Shrinkage and LASSO strategies in high dimensional Heteroscedastic models. Technical report, University of Windsor, Department of Mathematics and Statistics, Windsor, Canada.

Rabemananjara, R., Zakoïan, J.M., 1993. Threshold ARCH models and asymmetries in volatility. J. Appl. Econ. 8 (1), 31–49.

Ren, Y., Zhang, X., 2010. Subset selection for vector autoregressive processes via adaptive Lasso. Stat. Probab. Lett. 80 (23), 1705–1712.

Sensfuß, F., Ragwitz, M., Genoese, M., 2008. The merit-order effect: a detailed analysis of the price effect of renewable electricity generation on spot market prices in Germany. Energy Policy 36 (8), 3086–3094.

Slawski, M., Hein, M., et al., 2013. Non-negative least squares for high-dimensional linear models: consistency and sparse recovery without regularization. Electron. J. Stat. 7, 3004–3056.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B Stat Methodol. 58 (1), 267–288.

Wagener, J., Dette, H., 2012. Bridge estimators and the adaptive lasso under heteroscedasticity. Math. Meth. Stat. 21 (2), 109–126.

Wagener, J., Dette, H., 2013. The adaptive lasso in high-dimensional sparse heteroscedastic models. Math. Meth. Stat. 22 (2), 137–154.

Wang, H., Li, G., Tsai, C.-L., 2007. Regression coefficient and autoregressive order shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B Stat Methodol. 69 (1), 63–78.

Weron, R., 2006. Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach. John Wiley & Sons, Chichester, England.
Woo, C.-K., Horowitz, I., Moore, J., Pacheco, A., 2011. The impact of wind generation on the electricity spot-market price level and variance: the Texas experience. Energy Policy 39 (7), 3939–3944.
Würzburg, K., Labandeira, X., Linares, P., 2013. Renewable generation and electricity prices: taking stock and new evidence for Germany and Austria. Energy Econ. 40, 159–171.
Yan, X., Chowdhury, N.A., 2013. Mid-term electricity market clearing price forecasting: a hybrid LSSVM and ARMAX approach. Int. J. Electr. Power Energy Syst. 53, 20–26.