ORIGINAL PAPER

Efficient estimation of Markov regime-switching models: An application to electricity spot prices

Joanna Janczura · Rafał Weron

Abstract In this paper we discuss the calibration of models built on mean-reverting processes combined with Markov regime-switching (MRS). We propose a method that greatly reduces the computational burden induced by the introduction of independent regimes and perform a simulation study to test its efficiency. Our method allows for a 100 to over 1000 times faster calibration than in case of a competing approach utilizing probabilities of the last 10 observations. It is also more general and admits any value of γ in the base regime dynamics. Since the motivation for this research comes from a recent stream of literature in energy economics, we apply the new method to sample series of electricity spot prices from the German EEX and Australian NSW markets. The proposed MRS models fit these datasets well and replicate the major stylized facts of electricity spot price dynamics.

Keywords Markov regime-switching · Energy economics · Electricity spot price · EM algorithm · Independent regimes

1 Introduction

The underlying idea behind Markov regime-switching (MRS; or hidden Markov models—HMM) is to represent the observed stochastic behavior of a specific time
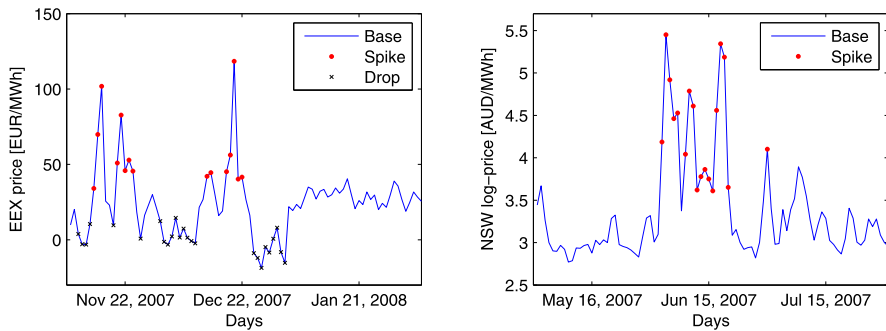
J. Janczura
Hugo Steinhaus Center, Institute of Mathematics and Computer Science, Wrocław University of Technology, Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland
e-mail: joanna.janczura@pwr.wroc.pl

R. Weron (✉)
Institute of Organization and Management, Wrocław University of Technology, Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland
e-mail: rafal.weron@pwr.wroc.pl

 Springer

Fig. 1 Deseasonalized mean daily electricity spot prices from the European Energy Exchange (EEX, Germany; *left*) and the New South Wales Electricity Market (NSW, Australia; *right*). Note, that the right panel uses a logarithmic scale to dampen the extreme spikiness of the Australian NSW prices, which can reach up to 10000 AUD/MWh during peak hours. The changes of dynamics (regime switches) are clearly visible in both cases. The prices classified as spikes or drops are denoted by *dots* or '×' (see Sect. 5 for model details)

series by two (or more) separate states or regimes with different underlying stochastic processes. The switching mechanism between the states is assumed to be an unobserved (latent) Markov chain. Such models have attracted a lot of attention in the recent years, not only in econometrics but also in other as diverse fields of science as population dynamics, speech recognition, river flow analysis or traffic modeling (Fink 2008; Hahn et al. 2009; Hamilton 2008; Luo and Mao 2007). This paper is motivated by yet another stream of literature—electricity spot price models in energy economics (Bierbrauer et al. 2007; De Jong 2006; Erlwein et al. 2010; Huisman and Mahieu 2003; Janczura and Weron 2010; Kanamura and Ōhashi 2008; Karakatsani and Bunn 2008; Kholodnyi 2005; Mari 2008; Mount et al. 2006; Weron 2009). MRS models have seen extensive use in this area due to their relative parsimony and the ability to capture the unique characteristics of electricity spot prices.

Recall, that electricity cannot be stored economically and requires immediate delivery. At the same time end-user demand shows high variability and strong weather and business cycle dependence. Effects like power plant outages, transmission grid (un)reliability and strategic bidding add complexity and randomness. The resulting spot prices exhibit strong seasonality on the annual, weekly and daily level, as well as, mean reversion, very high volatility and abrupt, short-lived and generally unanticipated extreme price spikes or drops (Benth et al. 2008a; Eydeland and Wolyniec 2003; Weron 2006). What is more, these spikes tend to cluster. Like in Fig. 1 where two sample spot price trajectories are plotted (for more evidence and discussions see Christensen et al. 2009; Janczura and Weron 2010). The latter feature renders the very popular class of jump–diffusion models impractical, as they cannot exhibit consecutive spikes with the frequency observed in market data.

In contrast, MRS models allow for consecutive spikes in a very natural way. Also the return of prices after a spike to the 'normal' regime is straightforward, as the regime-switching mechanism admits temporal changes of model dynamics. MRS models are also more versatile than the popular hidden Markov models (HMM; in the

strict sense, see Cappe et al. 2005), since they allow for temporary dependence within the regimes, in particular, for mean reversion. As the latter is a characteristic feature of electricity prices, it is important to have a model that captures this phenomenon. Indeed, the base regime is typically modeled by a mean-reverting diffusion (Huisman 2009), sometimes heteroskedastic (Janczura and Weron 2010). For the spike regime(s), on the other hand, a number of specifications have been suggested in the literature, ranging from mean-reverting diffusions to heavy tailed random variables.

Having selected the model class (i.e. MRS), the type of dependence between the regimes has to be defined. Dependent regimes with the same random noise process in all regimes (but different parameters; an approach dating back to the seminal work of Hamilton 1989) lead to computationally simpler models. On the other hand, independent regimes allow for a greater flexibility and admit qualitatively different dynamics in each regime. They seem to be a more natural choice for electricity spot price processes, which can exhibit a moderately volatile behavior in the base regime and a very volatile one in the spike regime, see Fig. 1 (note, that the right panel uses a logarithmic scale to dampen the extreme spikiness of the Australian NSW prices).

Once the electricity spot price model is fully specified we are left with the problem of calibrating it to market data. This challenging process is the focus of this paper. Due to the unobservable switching mechanism, estimation of MRS models requires inferring model parameters and state process values at the same time. The situation becomes more complicated when the individual regimes are independent from each other and at least one of them is mean-reverting. Then the temporal latency of the dynamics in the regimes has to be taken into account.

In this paper we propose a method that greatly reduces the computational burden induced by the introduction of independent regimes in MRS models. Instead of storing conditional probabilities for each of the possible state process paths, our method requires conditional probabilities for only one time-step. Since MRS models can be considered as generalizations of HMM (Cappe et al. 2005), this result can have far-reaching implications for many problems where HMM have been applied (see e.g. Mamon and Elliott 2007).

The paper is structured as follows. In Sect. 2 we define the MRS models used in this paper. Next, in Sect. 3 we describe the estimation procedure for parameter-switching models and introduce an approximation to avoid the computational burden in case of independent regimes. In Sect. 4 a simulation study to test the performance of the proposed method is summarized. Then, in Sect. 5 an application of the proposed approach to models of wholesale electricity prices is discussed. Finally, in Sect. 6 we conclude and provide suggestions for future work.

2 The models

Recall, that the switching mechanism between the states (or regimes) of a Markov regime-switching (MRS) model is assumed to be an unobserved (latent) Markov chain R_t. It is described by the transition matrix $\mathbf{P}$ containing the probabilities

$p_{ij} = P(R_{t+1} = j \mid R_t = i)$ of switching from regime i at time t to regime j at time $t+1$. For instance, for $i, j = \{1, 2\}$ we have

$$\mathbf{P} = (p_{ij}) = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} = \begin{pmatrix} 1 - p_{12} & p_{12} \\ p_{21} & 1 - p_{21} \end{pmatrix}. \tag{1}$$

Because of the Markov property the current state R_t at time t depends on the past only through the most recent value R_{t-1}.

In this paper we focus on two specifications of MRS models popular in the energy economics literature (see e.g. De Jong 2006; Janczura and Weron 2010; Mount et al. 2006). Both are based on a discretized version of the mean-reverting, heteroskedastic process given by the following SDE:

$$dX_t = (\alpha - \beta X_t)\, dt + \sigma |X_t|^\gamma\, dW_t. \tag{2}$$

Note, that the absolute value is needed if negative data is analyzed.

In the first specification only the model parameters change depending on the state process values, while in the second the individual regimes are driven by independent processes. More precisely, in the first case the observed process X_t is described by a parameter-switching times series of the form

$$X_t = \alpha_{R_t} + (1 - \beta_{R_t})X_{t-1} + \sigma_{R_t}|X_{t-1}|^{\gamma_{R_t}} \epsilon_t, \tag{3}$$

sharing the same set of random innovations in both regimes (ϵ_t's are assumed to be N(0, 1)-distributed). In the second one, X_t is defined as

$$X_t = \begin{cases} X_{t,1} & \text{if } R_t = 1, \\ X_{t,2} & \text{if } R_t = 2, \end{cases} \tag{4}$$

where at least one regime is given by

$$X_{t,i} = \alpha_i + (1 - \beta_i)X_{t-1,i} + \sigma_i|X_{t-1,i}|^{\gamma_i}\epsilon_{t,i}, \quad i = 1 \vee i = 2. \tag{5}$$

Note, that here we focus on a 2-regime model, but it is straightforward to generalize all the results of this paper to a model with three or more regimes.

3 Model calibration

Calibration of MRS models is not straightforward since the regimes are only latent and hence not directly observable. Hamilton (1990) introduced an application of the Expectation-Maximization (EM) algorithm of Dempster et al. (1977), where the whole set of parameters θ is estimated by an iterative two-step procedure. The algorithm was later refined by Kim (1994). In Sect. 3.1 we briefly describe the general estimation procedure and provide explicit formulas for the model defined by (3). Next, in Sect. 3.2 we discuss the computational problems induced by the introduction of independent regimes, see (4) and (5), and propose an efficient remedy.

3.1 Parameter-switching variant

The algorithm starts with an arbitrarily chosen vector of initial parameters $\theta^{(0)} = (\alpha_i^{(0)}, \beta_i^{(0)}, \sigma_i^{(0)}, \gamma_i^{(0)}, \mathbf{P}^{(0)}, \rho_i^{(0)})$, for $i = 1, 2$, where $\rho_i^{(0)} \equiv P(R_1 = i)$ and the other parameters are defined by (1), (3) and (5). In the first step of the iterative procedure (the E-step) inferences about the state process are derived. Since R_t is latent and not directly observable, only the expected values of the state process, given the observation vector $E(\mathbb{I}_{R_t=i} \mid x_1, x_2, \ldots, x_T; \theta)$, can be calculated. These expectations result in the so called 'smoothed inferences', i.e. the conditional probabilities $P(R_t = j \mid x_1, \ldots, x_T; \theta)$ for the process being in regime j at time t. Next, in the second step (the M-step) new maximum likelihood (ML) estimates of the parameter vector θ, based on the smoothed inferences obtained in the E-step, are calculated. Both steps are repeated until the (local) maximum of the likelihood function is reached. A detailed description of the algorithm is given below.

3.1.1 The E-step

Assume that $\theta^{(n)}$ is the parameter vector calculated in the M-step during the previous iteration. Let $\mathbf{x}_t = (x_1, x_2, \ldots, x_t)$. The E-part consists of the following steps (Kim 1994):

(i) *Filtering*: based on the Bayes rule for $t = 1, 2, \ldots, T$ iterate on equations:

$$P\left(R_t = i \mid \mathbf{x}_t; \theta^{(n)}\right) = \frac{P(R_t = i \mid \mathbf{x}_{t-1}; \theta^{(n)}) f(x_t \mid R_t = i; \mathbf{x}_{t-1}; \theta^{(n)})}{\sum_{i=1}^{2} P(R_t = i \mid \mathbf{x}_{t-1}; \theta^{(n)}) f(x_t \mid R_t = i; \mathbf{x}_{t-1}; \theta^{(n)})},$$

where $f(x_t \mid R_t = i; \mathbf{x}_{t-1}; \theta^{(n)})$ is the density of the underlying process at time t conditional that the process was in regime i ($i \in 1, 2$), and

$$P\left(R_{t+1} = i \mid \mathbf{x}_t; \theta^{(n)}\right) = \sum_{j=1}^{2} p_{ji}^{(n)} P\left(R_t = j \mid \mathbf{x}_t; \theta^{(n)}\right),$$

until $P(R_T = i \mid \mathbf{x}_T; \theta^{(n)})$ is calculated.

The starting point for the iteration is chosen as $P(R_1 = i \mid \mathbf{x}_0; \theta^{(n)}) = \rho_i^{(n)}$.

(ii) *Smoothing*: for $t = T - 1, T - 2, \ldots, 1$ iterate on

$$P\left(R_t = i \mid \mathbf{x}_T; \theta^{(n)}\right) = \sum_{j=1}^{2} \frac{P(R_t = i \mid \mathbf{x}_t; \theta^{(n)}) P(R_{t+1} = j \mid \mathbf{x}_T; \theta^{(n)}) p_{ij}^{(n)}}{P(R_{t+1} = j \mid \mathbf{x}_t; \theta^{(n)})}.$$

The above procedure requires derivation of $f(x_t \mid R_t = i; \mathbf{x}_{t-1}; \theta^{(n)})$ used in the filtering part (i). Observe, that the model definition (3) implies that X_t given X_{t-1} has a conditional Gaussian distribution with mean $\alpha_i + (1 - \beta_i)X_{t-1}$ and standard deviation $\sigma_i |X_{t-1}|^{\gamma_i}$ given by the following probability distribution function (pdf):

$$f\left(x_t \mid R_t = i; \mathbf{x}_{t-1}; \theta^{(n)}\right) = \frac{1}{\sqrt{2\pi} \sigma_i^{(n)} |x_{t-1}|^{\gamma_i^{(n)}}}$$

$$\cdot \exp\left\{-\frac{(x_t - (1 - \beta_i^{(n)})x_{t-1} - \alpha_i^{(n)})^2}{2(\sigma_i^{(n)})^2 |x_{t-1}|^{2\gamma_i^{(n)}}}\right\}. \quad (6)$$

3.1.2 The M-step

In the second step of the EM algorithm, new and more exact maximum likelihood (ML) estimates $\theta^{(n+1)}$ for all model parameters are calculated. Compared to standard ML estimation, where for a given pdf f the log-likelihood function $\sum_{t=1}^{T} \log f(x_t, \theta^{(n)})$ is maximized, here each component of this sum has to be weighted with the corresponding smoothed inference, since each observation x_t belongs to the ith regime with probability $P(R_t = i \mid \mathbf{x}_T; \theta^{(n)})$. In particular, for the model defined by (3) explicit formulas for the estimates can be derived by setting the partial derivatives of the (log-)likelihood function to zero. This leads to the following system of equations:

$$\hat{\alpha}_i = \frac{\sum_{t=2}^{T}[P(R_t = i \mid \mathbf{x}_T; \theta^{(n)})|x_{t-1}|^{-2\gamma_i}(x_t - (1 - \hat{\beta}_i)x_{t-1})]}{\sum_{t=2}^{T}[P(R_t = i \mid \mathbf{x}_T; \theta^{(n)})|x_{t-1}|^{-2\gamma_i}]}, \tag{7}$$

$$\hat{\beta}_i = \frac{\sum_{t=2}^{T}\{P(R_t = i \mid \mathbf{x}_T; \theta^{(n)})x_{t-1}|x_{t-1}|^{-2\gamma_i}B_1\}}{\sum_{t=2}^{T}[P(R_t = i \mid \mathbf{x}_T; \theta^{(n)})x_{t-1}|x_{t-1}|^{-2\gamma_i}B_2]},$$

$$B_1 = x_t - x_{t-1} - \frac{\sum_{t=2}^{T} P(R_t = i \mid \mathbf{x}_T; \theta^{(n)})|x_{t-1}|^{-2\gamma_i}(x_t - x_{t-1})}{\sum_{t=2}^{T} P(R_t = i \mid \mathbf{x}_T; \theta^{(n)})|x_{t-1}|^{-2\gamma_i}},$$

$$B_2 = \frac{\sum_{t=2}^{T} P(R_t = i \mid \mathbf{x}_T; \theta^{(n)})x_{t-1}|x_{t-1}|^{-2\gamma_i}}{\sum_{t=2}^{T} P(R_t = i \mid \mathbf{x}_T; \theta^{(n)})|x_{t-1}|^{-2\gamma_i}} - x_{t-1}, \tag{8}$$

$$\hat{\sigma}_i^2 = \frac{\sum_{t=2}^{T}\{P(R_t = i \mid \mathbf{x}_T; \theta^{(n)})|x_{t-1}|^{-2\gamma_i}(x_t - \hat{\alpha}_i - (1 - \hat{\beta}_i)x_{t-1})^2\}}{\sum_{t=2}^{T} P(R_t = i \mid \mathbf{x}_T; \theta^{(n)})}. \tag{9}$$

The fourth parameter, γ_i, requires numerical maximization of the (log-)likelihood function.

Finally, as in Hamilton (1990), we have $\rho_i^{(n+1)} = P(R_1 = i \mid \mathbf{x}_T; \theta^{(n)})$ and the transition probabilities are estimated according to the following formula (Kim 1994):

$$p_{ij}^{(n+1)} = \frac{\sum_{t=2}^{T} P(R_t = j, R_{t-1} = i \mid \mathbf{x}_T; \theta^{(n)})}{\sum_{t=2}^{T} P(R_{t-1} = i \mid \mathbf{x}_T; \theta^{(n)})}$$

$$= \frac{\sum_{t=2}^{T} P(R_t = j \mid \mathbf{x}_T; \theta^{(n)}) \frac{p_{ij}^{(n)} P(R_{t-1}=i|\mathbf{x}_{t-1};\theta^{(n)})}{P(R_t=j|\mathbf{x}_{t-1};\theta^{(n)})}}{\sum_{t=2}^{T} P(R_{t-1} = i \mid \mathbf{x}_T; \theta^{(n)})}, \tag{10}$$

where $p_{ij}^{(n)}$ is the transition probability from the previous iteration. All values obtained in the M-step are then used as a new parameter vector $\theta^{(n+1)} = (\hat{\alpha}_i, \hat{\beta}_i, \hat{\sigma}_i, \hat{\gamma}_i, \mathbf{P}^{(n+1)}, \rho_i^{(n+1)})$, $i = 1, 2$, in the next iteration of the E-step.

3.2 Independent regimes variant

In the parameter-switching model (3) the current value of the process depends on the last observation only, no matter which regime it originated from. This implies that for the calculation of the conditional pdf (6), used in the (i) part of the E-step recursions, the information from only one preceding time step is needed. Consequently, the EM
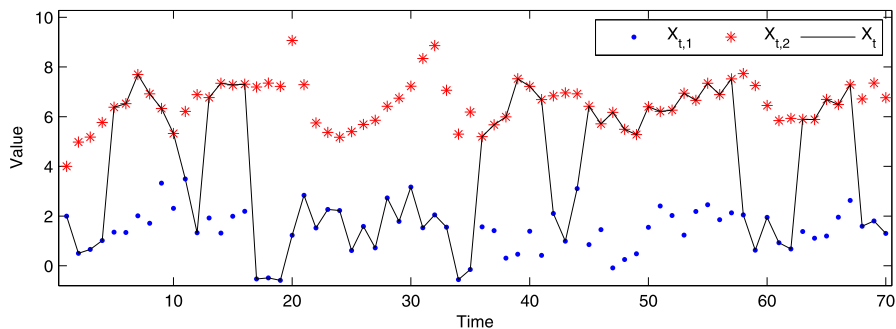
Fig. 2 A sample trajectory of a MRS model with independent regimes (*black solid line*) superimposed on the observable and latent values of the processes in both regimes. The simulation was performed for a model with the following parameters: $p_{11} = 0.9$, $p_{22} = 0.8$, $\alpha_1 = 1$, $\beta_1 = 0.7$, $\sigma_1^2 = 1$, $\gamma_1 = 0$, $\alpha_2 = 2$, $\beta_2 = 0.3$, $\sigma_2^2 = 0.01$, $\gamma_2 = 1$

algorithm requires storing conditional probabilities $P(R_t = i \mid \mathbf{x}_T)$ of one time step only, i.e. $2T$ values in total.

However, the estimation procedure complicates significantly, if the regimes are independent from each other. Observe, that the values of the mean-reverting regime become latent when the process is in the other state (see Fig. 2 for an illustration). This makes the distribution of X_t dependent on the whole history $(x_1, x_2, \ldots, x_{t-1})$ of the process. As a consequence all possible paths of the state process $(R_1, R_2, \ldots, R_t)$ should be considered in the estimation procedure, implying that $f(x_t \mid R_t = i, R_{t-1} \neq i, \ldots, R_{t-j} \neq i, R_{t-j-1} = i; \mathbf{x}_{t-1}; \theta^{(n)})$ and the whole set of probabilities $P(R_t = i_t, R_{t-1} = i_{t-1}, \ldots, R_{t-j} = i_{t-j} \mid \mathbf{x}_{t-1}; \theta^{(n)})$ should be used in the E-step. Obviously, this leads to a high computational complexity, as the number of possible state process realizations is equal to 2^T and increases rapidly with sample size. To be more precise, the total number of probabilities required by the EM algorithm to be stored in computer memory is equal to $2(2^{T+1} - 1)$. Assuming that each probability is stored as a double precision floating-point number (8 bytes), estimating parameters from a sample of $T = 30$ observations would require 32 gigabytes of memory! For samples of typical size (a few hundred to a few thousand observations) this is clearly impossible with today's computers.

As a feasible solution to this problem Huisman and de Jong (2002) suggested to use probabilities of the last 10 observations. Apart from the fact that such an approximation still is computationally intensive (requires storing $2\{2^{10}(T - 9) - 1\}$ probabilities in computer memory), it can be used only if the probability of more than 10 consecutive observations from the second (spike) regime is negligible. In Sect. 4 we will perform a simulation study to see how limiting is this theoretical assumption in practice.

Instead, in this article we suggest to replace the latent variables x_{t-1} in formula (6) with their expectations $\tilde{x}_{t-1} = E(X_{t-1} \mid \mathbf{x}_{t-1}; \theta^{(n)})$ based on the whole information available at time $t - 1$. A similar approach was used by Gray (1996) in the context of regime-switching GARCH models to avoid the problem of the conditional standard deviation path dependence. Now, the estimation procedure described in Sect. 3.1 can be applied with the following approximation of the pdf:

$$f\left(x_t \mid R_t = i; \mathbf{x}_{t-1}; \theta^{(n)}\right) = \frac{1}{\sqrt{2\pi}\sigma_i^{(n)}|\tilde{x}_{t-1,i}|^{\gamma_i^{(n)}}}$$

$$\cdot \exp\left\{-\frac{(x_t - (1 - \beta_i^{(n)})\tilde{x}_{t-1,i} - \alpha_i^{(n)})^2}{2(\sigma_i^{(n)})^2|\tilde{x}_{t-1,i}|^{2\gamma_i^{(n)}}}\right\}, \quad (11)$$

where $\tilde{x}_{t,i}$ denotes the expected value of the ith regime at time t, that is $E(X_{t,i} \mid \mathbf{x}_t; \theta^{(n)})$. Note, that compared to formula (6) for the parameter-switching variant, the observed value of the process x_{t-1} is now replaced by the expected value $\tilde{x}_{t-1,i}$ of the ith regime at time $t-1$. The expected values $\tilde{x}_{t,i} = E(X_{t,i} \mid \mathbf{x}_t; \theta^{(n)})$ can be computed using the following recursive formula (for the derivation see the Appendix):

$$E\left(X_{t,i} \mid \mathbf{x}_t; \theta^{(n)}\right) = P\left(R_t = i \mid \mathbf{x}_t; \theta^{(n)}\right)x_t + P\left(R_t \neq i \mid \mathbf{x}_t; \theta^{(n)}\right)$$
$$\cdot \left\{\alpha_i^{(n)} + \left(1 - \beta_i^{(n)}\right)E\left(X_{t-1,i} \mid \mathbf{x}_{t-1}; \theta^{(n)}\right)\right\}. \quad (12)$$

It is interesting to note that

$$E\left(X_{t,i} \mid \mathbf{x}_t; \theta^{(n)}\right) = \sum_{k=0}^{t-1} x_{t-k}\left(1 - \beta_i^{(n)}\right)^k P\left(R_{t-k} = i \mid \mathbf{x}_{t-k}; \theta^{(n)}\right)$$

$$\cdot \prod_{j=1}^{k} P\left(R_{t-j+1} \neq i \mid \mathbf{x}_{t-j+1}; \theta^{(n)}\right)$$

$$+ \alpha_i^{(n)} \sum_{k=0}^{t-1}\left(1 - \beta_i^{(n)}\right)^k \prod_{j=0}^{k} P\left(R_{t-j+1} \neq i \mid \mathbf{x}_{t-j+1}; \theta^{(n)}\right).$$

Hence, the expected value $E(X_{t,i} \mid \mathbf{x}_t; \theta^{(n)})$ is a linear combination of the observed vector $\mathbf{x}_t$ and the probabilities $P(R_j = i \mid \mathbf{x}_j; \theta^{(n)})$ calculated during the estimation procedure (see the filtering part of the E-step). This observation shows that using $\tilde{x}_{t-1,i} = E(X_{t-1,i} \mid \mathbf{x}_{t-1}; \theta^{(n)})$ in formula (11) instead of x_{t-1}, as in formula (6) for the parameter-switching variant, the computational complexity of the E-step is greatly reduced. In particular, in the proposed modification of the EM algorithm the total number of probabilities stored in computer memory is only $4T$. This means that for a sample of $T = 30$ observations only 1 kilobyte of memory is required, compared to 335 kilobytes in the approach utilizing probabilities of the last 10 observations and 32 gigabytes in the standard EM algorithm.

The estimation procedure described in this section can be applied to models in which at least one regime is described by the mean-reverting process given by (5). The independent regimes specification is commonly used in the electricity price modeling literature (for a recent review see Janczura and Weron 2010). It is often assumed that one regime follows a mean-reverting process, while the values in the other regime(s) are independent random variables from a specified distribution. The estimation steps are then as described above, with the exception that the M-step is now dependent on the choice of the distribution in the other regime(s). Finally, note that in MRS models the likelihood function should be weighted with the corresponding probability, analogously as in the derivation of estimates (7)–(9) in the parameter-switching variant.

4 Simulation study

In order to test the performance of the estimation method proposed in Sect. 3.2, we provide a simulation study. For each of the following three MRS model types we generate 1000 sample trajectories:

– MR: with parameter-switching mean-reverting regimes, see (3).
– IMR: with independent mean-reverting processes in both regimes, see (5).
– IMR-G: with a mean-reverting process in the first regime and independent $N(\alpha_2, \sigma_2^2)$-distributed random variables in the second regime.

The IMR model is simulated with probabilities of staying in the same regime equal to $p_{11} = 0.9$ and $p_{22} = 0.8$ for the first and the second regime, respectively. With such a choice of the transition matrix we can expect to see many consecutive observations in each regime. Indeed, the probability of 10 consecutive observations from the first regime is equal to 0.35 and even for 40 consecutive observations that probability is still higher than 0.01. Obviously, such a model cannot be estimated based on the information about only a few prevailing observations.

For each sample trajectory we apply one of the estimation procedures described in Sect. 3. Then, we calculate the means, standard deviations and 95% confidence intervals of the parameter estimates. The values obtained for trajectories consisting of 10000 observations are given in Table 1. All sample means are close to the true parameters with a deviation of no more than 0.03 (in absolute terms). In fact, in most cases the deviation is significantly lower. Moreover, all parameter values are within the obtained 95% confidence intervals. Also the standard deviation of the estimates is quite low and, except for γ_2 and α_2 in the IMR model, does not exceed 0.04.

Next, we check how the proposed method works for different sample sizes. We generate MRS model trajectories with 100, 500, 1000, 2000, 5000, and 10000 observations. The obtained means and standard deviations are given in Table 2 (MR model), Table 3 (IMR model) and Table 4 (IMR-G model). The respective confidence intervals are plotted in Figs. 3, 4 and 5. As expected, the standard deviations, as well as, the width of the confidence intervals decrease with increasing sample size. Looking at the means, in most cases a sample of 1000 (or even 500 for the MR and IMR-G models) observations yields satisfactory results, as the deviation does not exceed 0.03 (in absolute terms). Especially for the IMR-G model the results are very satisfactory. This is important in view of the fact that a variant of this model is used in Sect. 5 for modeling electricity spot prices.

Finally, we compare the convergence of the proposed algorithm (in what follows called method I) with the approach utilizing probabilities of the last 10 observations (method II; as proposed by Huisman and de Jong 2002). Observe, that since the conditional distribution $f(x_t \mid x_{t-k})$ of the process defined by (5) is not known for a general value of γ if $k > 1$, the latter method cannot be used in this case. Therefore, we limit the comparison of the two estimation methods to specifications with $\gamma = 0$. For 100 simulated trajectories of the IMR-G model we calculate the mean absolute errors (MAE) of the parameter estimates obtained from both approaches. We consider five sample sizes (100, 500, 1000, 2000 and 5000) and two sets of parameters: one with a low probability of staying in the Gaussian regime (i.e. with $p_{22} = 0.2$, so that

Table 1 Means, 95% confidence intervals (CI$_l$,CI$_u$) and standard deviations (Std) of parameter estimates obtained from 1000 simulated trajectories of 10000 observations each, for the three studied MRS model types: MR, IMR, and IMR-G

	α_1	β_1	σ_1^2	γ_1	α_2	β_2	σ_2^2	γ_2	p_{11}	p_{22}
MR										
True	1.0000	0.7000	1.0000	0.0000	2.0000	0.3000	0.0100	1.0000	0.5000	0.5000
Mean	1.0006	0.7001	1.0004	−0.0002	2.0000	0.3000	0.0100	1.0010	0.4998	0.4993
CI$_l$	0.9493	0.6842	0.9504	−0.0223	1.9997	0.2974	0.0094	0.9752	0.4854	0.4856
CI$_u$	1.0587	0.7157	1.0543	0.0223	2.0003	0.3026	0.0106	1.0251	0.5134	0.5128
Std	0.0332	0.0095	0.0316	0.0137	0.0002	0.0016	0.0004	0.0152	0.0083	0.0081
IMR										
True	1.0000	0.7000	1.0000	0.0000	2.0000	0.3000	0.0100	1.0000	0.9000	0.8000
Mean	0.9974	0.6973	1.0135	0.0029	1.9702	0.2957	0.0128	1.0010	0.9004	0.8003
CI$_l$	0.9633	0.6778	0.9836	−0.0126	1.8347	0.2753	0.0026	0.8219	0.8941	0.7892
CI$_u$	1.0339	0.7189	1.0435	0.0176	2.1145	0.3181	0.0496	1.1716	0.9066	0.8113
Std	0.0216	0.0126	0.0184	0.0094	0.0857	0.0131	0.0055	0.1051	0.0038	0.0068
IMR-G										
True	1.0000	0.7000	0.5000	0.5000	7.0000		0.5000		0.8000	0.000
Mean	0.9999	0.7009	0.5074	0.5036	6.9959		0.5063		0.7999	0.2018
CI$_l$	0.9893	0.6888	0.4924	0.4865	6.9689		0.4801		0.7928	0.1876
CI$_u$	1.0113	0.7136	0.5220	0.5209	7.0218		0.5348		0.8071	0.2163
Std	0.0067	0.0075	0.0087	0.0105	0.0160		0.0165		0.0044	0.0089

the probability of more than 10 consecutive observations from the Gaussian regime is less than 10^{-7} and, hence, can be neglected) and a second one with $p_{22} = 0.7$. The results are summarized in Table 5.

The mean absolute errors obtained for the first parameter set (with $p_{22} = 0.2$) are comparable for both methods. However, if the probability of staying in the Gaussian regime for more than 10 consecutive observations is not negligible, method I apparently outperforms method II. The difference is especially noticeable for the Gaussian regime parameters (α_2 and σ_2^2) and the transition probabilities. Observe, that for samples of 5000 observations the MAE obtained using method II are almost five times larger than the ones resulting from using method I. Moreover, while the errors decrease with increasing sample size when using method I, this is not observed for method II, e.g. for the estimates of α_2 the MAE values are 0.1005, 0.1045, 0.0876 and 0.1036 for samples of 500, 1000, 2000 and 5000 observations, respectively.

This simple simulation study shows the appealing efficiency of the estimation algorithm proposed in this article (method I), when compared to its competitor (method II). It also makes clear that method II can be used only if the probability of more than 10 consecutive observations from the spike (here Gaussian) regime can be ignored. Furthermore, compared to method I, method II is much more computationally demanding. In the last column of Table 5 we provide the mean (over 100

Table 2 Means and standard deviations (Std), over 1000 simulated trajectories, of parameter estimates in the MR model calculated for different sample sizes

	α_1	β_1	σ_1^2	γ_1	α_2	β_2	σ_2^2	γ_2	p_{11}	p_{22}
True	1.0000	0.7000	1.0000	0.0000	2.0000	0.3000	0.0100	1.0000	0.5000	0.5000
Size	Mean									
100	0.9904	0.7040	0.8863	0.0455	1.9990	0.3011	0.0088	1.1214	0.4895	0.4962
500	1.0027	0.7000	0.9818	0.0069	2.0000	0.2999	0.0096	1.0201	0.4986	0.4986
1000	1.0034	0.7014	0.9915	0.0021	2.0002	0.3002	0.0097	1.0151	0.4990	0.4989
2000	1.0021	0.7009	0.9966	0.0026	2.0000	0.3000	0.0099	1.0045	0.5002	0.4990
5000	1.0003	0.7003	0.9958	0.0005	2.0000	0.3000	0.0099	1.0035	0.4998	0.5000
10000	1.0006	0.7001	1.0004	−0.0002	2.0000	0.3000	0.0100	1.0010	0.4998	0.4993
Size	Std									
100	0.3746	0.1111	0.3903	0.2280	0.0279	0.0219	0.0056	0.2797	0.0864	0.0865
500	0.1563	0.0461	0.1535	0.0703	0.0042	0.0077	0.0018	0.0766	0.0350	0.0362
1000	0.1094	0.0311	0.1035	0.0453	0.0020	0.0051	0.0013	0.0518	0.0263	0.0256
2000	0.0719	0.0214	0.0719	0.0298	0.0010	0.0035	0.0008	0.0335	0.0179	0.0180
5000	0.0474	0.0139	0.0455	0.0196	0.0004	0.0023	0.0005	0.0213	0.0115	0.0115
10000	0.0332	0.0095	0.0316	0.0137	0.0002	0.0016	0.0004	0.0152	0.0083	0.0081

Table 3 Means and standard deviations (Std), over 1000 simulated trajectories, of parameter estimates in the IMR model calculated for different sample sizes

	α_1	β_1	σ_1^2	γ_1	α_2	β_2	σ_2^2	γ_2	p_{11}	p_{22}
True	1.0000	0.7000	1.0000	0.0000	2.0000	0.3000	0.0100	1.0000	0.9000	0.8000
Size	Mean									
100	1.0221	0.7280	0.9765	0.0298	2.6473	0.3982	22647	1.0580	0.8951	0.7798
500	1.0002	0.7026	1.0109	0.0061	2.0368	0.3057	0.5569	0.9822	0.8997	0.7956
1000	1.0037	0.7025	1.0070	0.0075	2.0319	0.3052	0.0391	0.9995	0.9004	0.7981
2000	0.9992	0.6987	1.0130	0.0024	1.9781	0.2969	0.0200	1.0046	0.9006	0.7993
5000	1.0001	0.6988	1.0128	0.0044	1.9615	0.2944	0.0139	1.0059	0.9004	0.8001
10000	0.9974	0.6973	1.0135	0.0029	1.9702	0.2957	0.0128	1.0010	0.9004	0.8003
Size	Std									
100	0.2254	0.1358	0.1975	0.1341	1.3062	0.1986	71587	2.1783	0.0385	0.0779
500	0.0954	0.0558	0.0869	0.0487	0.4207	0.0641	9.0870	0.5930	0.0166	0.0318
1000	0.0668	0.0403	0.0601	0.0312	0.2917	0.0444	0.1958	0.3687	0.0116	0.0218
2000	0.0473	0.0275	0.0408	0.0211	0.2006	0.0306	0.0620	0.2397	0.0083	0.0156
5000	0.0296	0.0176	0.0263	0.0135	0.1213	0.0184	0.0093	0.1606	0.0052	0.0098
10000	0.0216	0.0126	0.0184	0.0094	0.0857	0.0131	0.0055	0.1051	0.0038	0.0068

simulations) estimation times obtained using both approaches. Observe the striking difference between the two methods. Method I was found to be 100 to over 1000 times faster than method II (1.13 s vs. 1340.04 s)!

Table 4 Means and standard deviations (Std), over 1000 simulated trajectories, of parameter estimates in the IMR-G model calculated for different sample sizes

	α_1	β_1	σ_1^2	γ_1	α_2	σ_2^2	p_{11}	p_{22}
True	1.0000	0.7000	0.5000	0.5000	7.0000	0.5000	0.8000	0.2000
Size	Mean							
100	1.0017	0.7002	0.5064	0.5876	7.0084	0.4732	0.7977	0.1889
500	1.0045	0.7048	0.5058	0.5221	7.0007	0.5068	0.8011	0.2008
1000	0.9997	0.7004	0.5066	0.5137	6.9941	0.5066	0.7995	0.2012
2000	1.0007	0.7012	0.5086	0.5071	6.9971	0.5038	0.8001	0.2020
5000	1.0000	0.7000	0.5072	0.5055	6.9955	0.5068	0.8001	0.2019
10000	0.9999	0.7009	0.5074	0.5036	6.9959	0.5063	0.7999	0.2018
Size	Std							
100	0.1021	0.0927	0.1062	0.1622	0.1617	0.1730	0.0442	0.0910
500	0.0380	0.0370	0.0403	0.0563	0.0755	0.0786	0.0194	0.0395
1000	0.0252	0.0257	0.0273	0.0374	0.0510	0.0545	0.0147	0.0277
2000	0.0165	0.0174	0.0189	0.0251	0.0362	0.0377	0.0100	0.0192
5000	0.0098	0.0113	0.0121	0.0153	0.0232	0.0235	0.0065	0.0127
10000	0.0067	0.0075	0.0087	0.0105	0.0160	0.0165	0.0044	0.0089

5 Application to electricity spot prices

In this study we present how the techniques introduced in Sect. 3 can be used to efficiently calibrate MRS models to electricity spot prices. We use mean daily (baseload) spot prices from two major power markets: the European Energy Exchange (EEX; Germany) and the New South Wales Electricity Market (NSW; being part of the National Electricity Market in Australia). Using baseload data is quite common in the energy economics literature, partly due to the fact that baseload is the most common underlying instrument for energy derivatives. The EEX sample totals 1820 daily observations (or 260 full weeks) and covers the roughly 5-year period January 2, 2006–December 26, 2010. The NSW sample totals 1722 daily observations (246 full weeks) and covers the period January 2, 2006–September 19, 2010. Recall that NSW is an 'energy only market', where the wholesale electricity price provides compensation for both variable and fixed costs (Weron 2006). As a result, the observed prices are extremely spiky—they can reach up to 10000 AUD/MWh during peak hours and well over 1000 AUD/MWh in mean daily prices. On the other hand, a different generation stack with much less wind generation than in the EEX market yields no negative (or close to zero) prices. Consequently, in what follows we analyze NSW log-prices, i.e. natural logarithms of spot prices, which are more prone to modeling with the models studied in this article than the prices themselves.

When modeling electricity spot prices we have to bear in mind that electricity is a very specific commodity. Both electricity demand and (to some extent) supply exhibit seasonal fluctuations. They mostly arise due to changing climate conditions, like temperature and the number of daylight hours. These seasonal fluctuations translate into seasonal behavior of electricity prices, and spot prices in particular. In the mid-
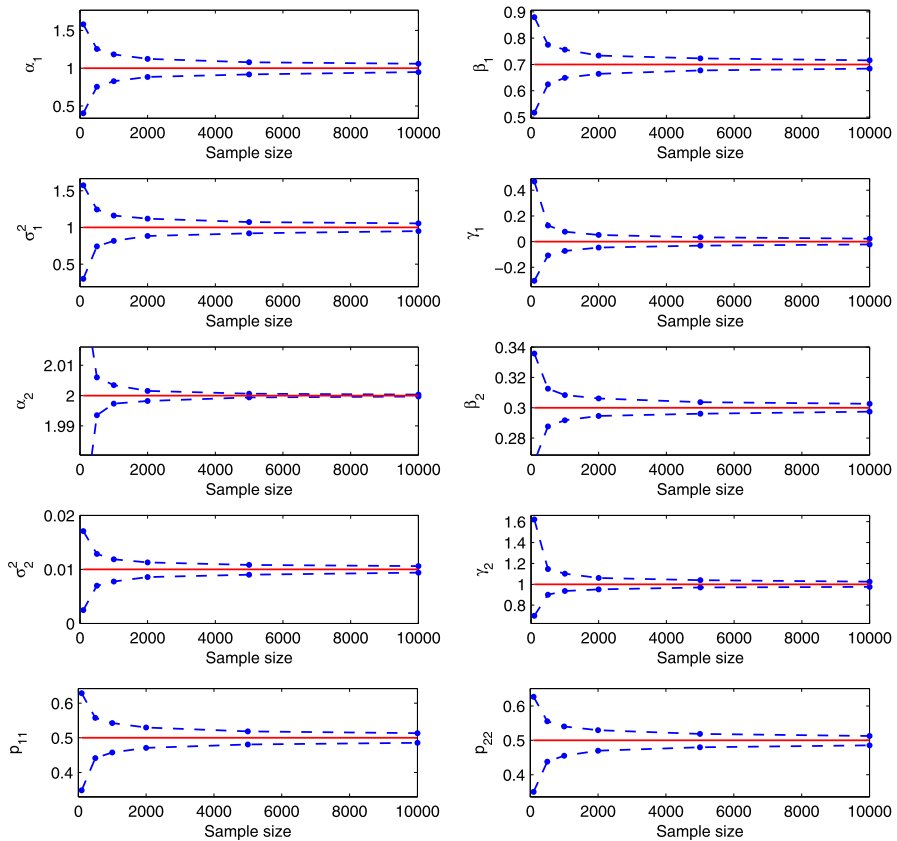
Fig. 3 (Color online) 95% confidence intervals of parameter estimates in the MRS model with parameter-switching mean-reverting regimes (MR; see Table 1 for parameter details). The true parameter values are given by the *solid red lines*

and long-term also the fuel price levels (of natural gas, oil, coal) influence electricity prices. However, not wanting to focus the paper on modeling the fuel stack/bid stack/electricity spot price relationships, we use a single non-parametric long-term seasonal component (LTSC) to represent the long-term non-periodic fuel price levels, the changing climate/consumption conditions throughout the years and strategic bidding practices. An empirical justification for such an approach can be found, for instance, in Janczura and Weron (2010); see also Eydeland and Wolyniec (2003) and Karakatsani and Bunn (2010) for discussions on fundamental and behavioral drivers of electricity prices.

We assume that the electricity spot price (or log-price for the NSW power market), P_t, can be represented by a sum of two independent parts: a predictable (seasonal) component f_t and a stochastic component X_t, i.e. $P_t = f_t + X_t$. Further, we let f_t be composed of a weekly periodic part, s_t, and a LTSC, T_t. The deseasonalization is then conducted in three steps. First, the long term trend T_t is estimated from daily spot prices P_t using a wavelet filtering–smoothing technique (for details see Trück
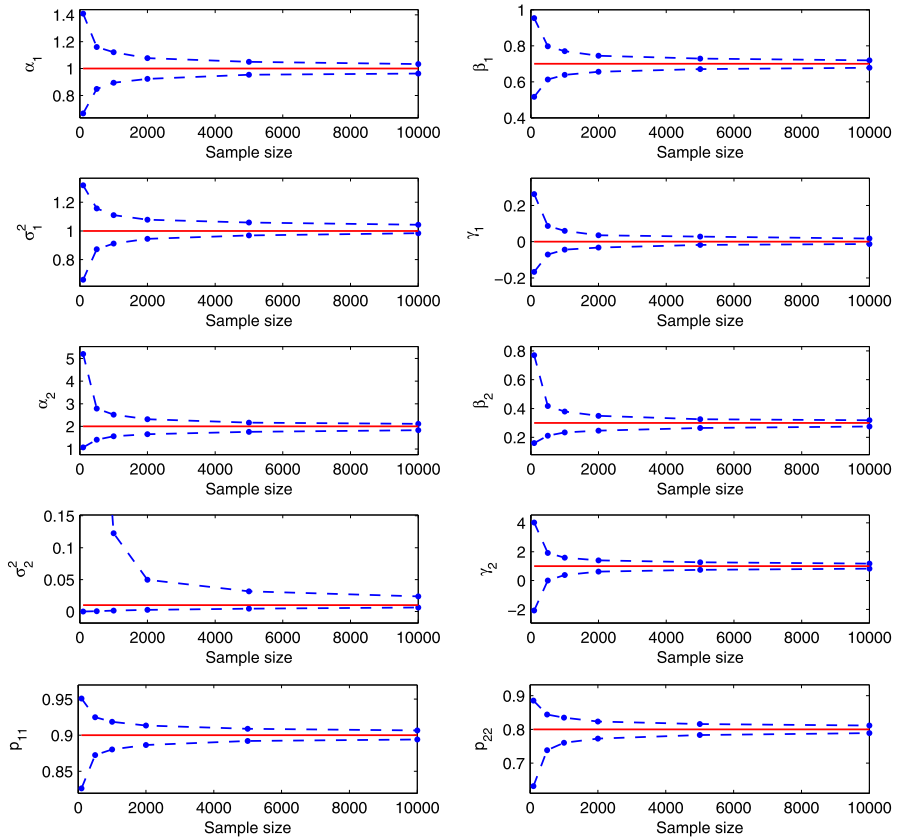
Fig. 4 (Color online) 95% confidence intervals of parameter estimates in the MRS model with independent mean-reverting regimes (IMR; see Table 1 for parameter details). The true parameter values are given by the *solid red lines*

et al. 2007; Weron 2006). This procedure, also known as low pass filtering, yields a traditional linear smoother. Here we use the S_6 approximation, which roughly corresponds to bi-monthly ($2^6 = 64$ days) smoothing. The estimated long term seasonal components are plotted in Fig. 6.

Although forecasting is not the focus of this article, let us briefly comment on forecasting the trend-seasonal patterns. Predicting the LTSC beyond the next few weeks is a difficult task. And wavelets are not very helpful in this context. Unlike sines and cosines, individual wavelet functions are quite localized in time or (more generally) in space. For the lower detail levels of the wavelet decomposition, which are of high frequency and oscillatory in nature, a trigonometric (Yousefi et al. 2005) or an ARIMA (Conejo et al. 2005) fit can be used. However, for the much smoother approximation level (e.g. the S_6 approximation in our case) and the higher detail levels rather a spline fit (Yousefi et al. 2005) or a polynomial extrapolation (Wong et al. 2003) is applied to extend the signal. Naturally, the choice of the extrapolation method and the range of values used for calibration is subjective and, hence, the LTSC
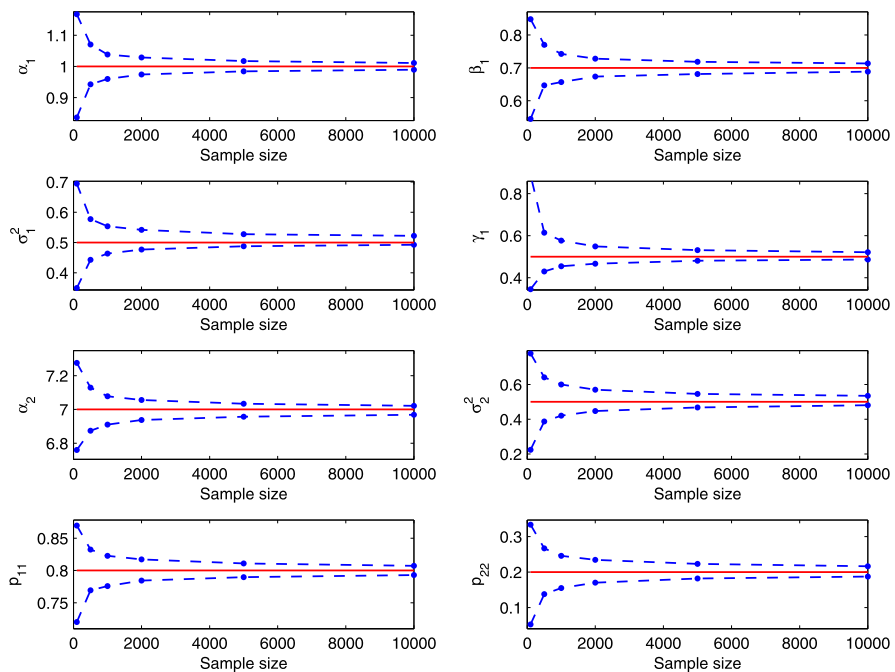
Fig. 5 (Color online) 95% confidence intervals of parameter estimates in the MRS model with a mean-reverting regime combined with independent Gaussian random variables (IMR-G; see Table 1 for parameter details). The true parameter values are given by the *solid red lines*

forecasts will most likely diverge for different methods. As mentioned by Janczura and Weron (2010), a potentially promising, alternative approach would be to use forward looking information, like smoothed forward curves (Benth et al. 2007; Borak and Weron 2008). The information carried by forward prices provides insights as to the future evolution of spot prices. However, forward prices also include the risk premium (Benth et al. 2008b; Weron 2008), which should somehow be separated from the spot price forecast for it to be useful.

The price series without the LTSC is obtained by subtracting the S_6 approximation from P_t. Next, the weekly periodicity s_t is removed by subtracting the 'average week' calculated as the median of prices corresponding to each day of the week (the median is used instead of the mean due to its robustness against outliers—the extremely spiky prices, especially in NSW data). Finally, the deseasonalized prices, i.e. $P_t - T_t - s_t$, are shifted so that the mean of the new process is the same as the mean of P_t. The resulting deseasonalized time series $X_t = P_t - T_t - s_t$ can be seen in Figs. 7 and 8.

The second well known feature of electricity prices are the sudden, unexpected price changes, known as spikes or jumps. The 'spiky' nature of spot prices is the effect of non-storability of electricity. Electricity to be delivered at a specific hour cannot be substituted for electricity available shortly after or before. Extreme load fluctuations—caused by severe weather conditions often in combination with generation outages or transmission failures—can lead to price spikes. On the other hand, an oversupply—due to a sudden drop in demand and technical limitations of an instant

Table 5 Mean absolute errors (MAE) of the parameter estimates in the IMR-G model obtained from two estimation methods: the algorithm proposed in this paper (method I) and the approach utilizing probabilities of the last 10 observations (method II). The errors were computed over 100 simulations for each sample size and each set of parameters. The true (simulated) parameter values are given in italics. Additionally, the mean estimation times (in seconds) are provided in the last column

Sample size	Method	α_1	β_1	σ_1^2	α_2	σ_2^2	p_{11}	p_{22}	Time [s]
		1.00	*0.70*	*0.50*	*7.00*	*0.50*	*0.80*	*0.20*	
100	I	0.1469	0.0954	0.0716	0.1291	0.1187	0.0331	0.0674	0.02
	II	0.1443	0.0929	0.0720	0.1287	0.1190	0.0332	0.0674	1.53
500	I	0.0573	0.0362	0.0293	0.0594	0.0606	0.0166	0.0375	0.07
	II	0.0592	0.0371	0.0274	0.0594	0.0606	0.0166	0.0375	8.68
1000	I	0.0398	0.0257	0.0199	0.0463	0.0455	0.0119	0.0247	0.14
	II	0.0378	0.0249	0.0192	0.0463	0.0455	0.0119	0.0247	19.67
2000	I	0.0318	0.0202	0.0151	0.0286	0.0259	0.0078	0.0149	0.28
	II	0.0309	0.0195	0.0139	0.0307	0.0266	0.0078	0.0152	35.71
5000	I	0.0219	0.0133	0.0122	0.0168	0.0192	0.0051	0.0098	0.67
	II	0.0342	0.0154	0.0565	0.0199	0.0238	0.0069	0.0111	109.25
		1.00	*0.60*	*1.00*	*8.00*	*1.00*	*0.90*	*0.70*	
100	I	0.1627	0.0803	0.1314	0.1708	0.3024	0.0275	0.0900	0.03
	II	0.1662	0.0832	0.1342	0.1802	0.2867	0.0319	0.0961	2.81
500	I	0.0788	0.0409	0.0738	0.0736	0.1152	0.0144	0.0298	0.12
	II	0.0777	0.0392	0.0719	0.1005	0.1503	0.0193	0.0424	32.70
1000	I	0.0575	0.0296	0.0409	0.0591	0.0731	0.0090	0.0215	0.22
	II	0.0586	0.0297	0.0394	0.1045	0.1516	0.0298	0.0417	135.42
2000	I	0.0385	0.0205	0.0365	0.0377	0.0552	0.0063	0.0162	0.45
	II	0.0435	0.0229	0.0802	0.0876	0.1332	0.0170	0.0405	403.95
5000	I	0.0262	0.0149	0.0249	0.0218	0.0332	0.0041	0.0097	1.13
	II	0.0309	0.0163	0.0758	0.1036	0.1404	0.0189	0.0445	1340.04

shut-down of a generator—can cause price drops. Further, electricity spot prices are in general regarded to be mean-reverting and exhibit the so called 'inverse leverage effect', meaning that the positive shocks increase volatility more than the negative shocks. Knittel and Roberts (2005) attributed this phenomenon to the fact that a positive shock to electricity prices can be treated as an unexpected positive demand shock. Therefore, as a result of convex marginal costs, positive demand shocks have a larger impact on price changes relative to negative shocks.

Motivated by these features of electricity spot prices we let the stochastic component X_t be driven by a Markov regime-switching model with three independent states:

$$X_t = \begin{cases} X_{t,1} & \text{if } R_t = 1, \\ X_{t,2} & \text{if } R_t = 2, \\ X_{t,3} & \text{if } R_t = 3. \end{cases} \tag{13}$$
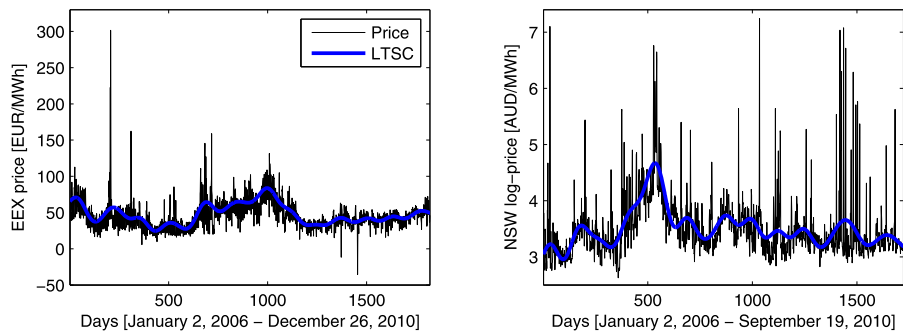
Fig. 6 (Color online) Mean daily spot EEX prices (*left panel*) and NSW log-prices (*right panel*) and the estimated long-term seasonal components (LTSC; *thick blue lines*)
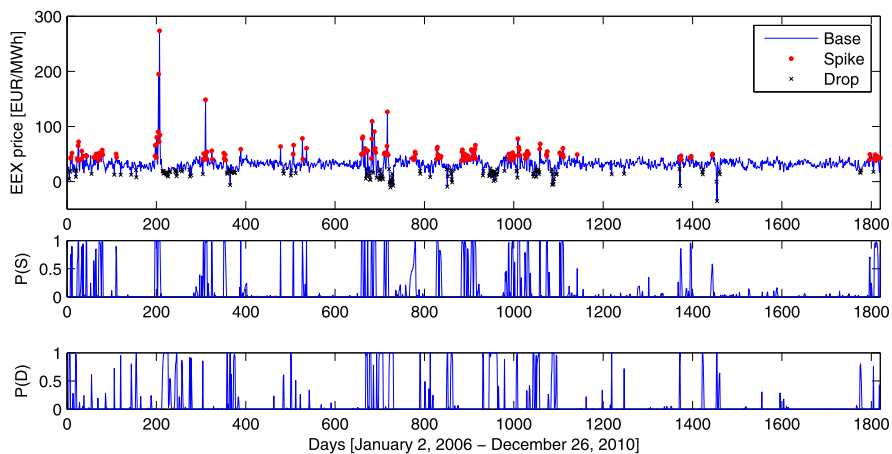
Fig. 7 Calibration results of a MRS model with three independent regimes fitted to the deseasonalized EEX prices. The *lower panels* display the conditional probabilities $P(S) = P(R_t = 2 \mid x_1, x_2, \ldots, x_T)$ and $P(D) = P(R_t = 3 \mid x_1, x_2, \ldots, x_T)$ of being in the spike or drop regime, respectively. The prices classified as spikes or drops, i.e. with $P(S) > 0.5$ or $P(D) > 0.5$, are denoted by *dots* or '$\times$' in the *upper panel*

The first (base) regime describes the 'normal' price behavior and is given by the mean-reverting, heteroskedastic process of the form:

$$X_{t,1} = \alpha_1 + (1 - \beta_1)X_{t-1,1} + \sigma_1 |X_{t-1,1}|^{\gamma_1} \epsilon_t, \tag{14}$$

where ϵ_t is the standard Gaussian noise. The second regime represents the sudden price jumps (spikes) caused by unexpected supply shortages and is given by i.i.d. random variables from the shifted log-normal distribution:

$$\log(X_{t,2} - X(q_2)) \sim N(\alpha_2, \sigma_2^2), \quad X_{t,2} > X(q_2). \tag{15}$$

Finally, the third regime (responsible for the sudden price drops) is governed by the shifted 'inverse log-normal' law:

$$\log(-X_{t,3} + X(q_3)) \sim N(\alpha_3, \sigma_3^2), \quad X_{t,3} < X(q_3). \tag{16}$$
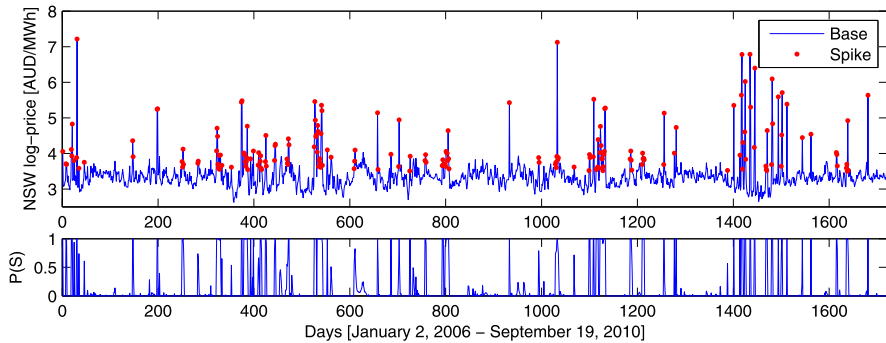
Fig. 8 Calibration results of a MRS model with two independent regimes fitted to the deseasonalized NSW log-prices. The *lower panel* displays the conditional probability $P(S)$ of being in the spike regime. The prices classified as spikes, i.e. with $P(S) > 0.5$, are denoted by *dots* in the *upper panel*

Table 6 Calibration results for MRS models with (two or three) independent regimes fitted to the deseasonalized EEX prices and NSW log-prices

	Parameters								Probabilities		
	α_1	β_1	σ_1^2	γ_1	α_2	σ_2^2	α_3	σ_3^2	p_{11}	p_{22}	p_{33}
EEX	14.08	0.44	7.23	0.18	2.34	0.87	2.49	0.33	0.9270	0.6420	0.6603
NSW	0.67	0.20	4e-4	1.60	−0.99	1.40	–	–	0.9539	0.6390	–

It turns out that, unlike EEX prices, NSW prices do not exhibit significant price drops, even on the log-scale. A third regime is not needed to adequately model the dynamics. Hence, in what follows we fit a 2-regime model (without a price drop regime) to the deseasonalized NSW log-prices.

In the above formulas $X(q_i)$ denotes the q_i-quantile, $q_i \in (0, 1)$, of the dataset. Generally the choice of q_i is arbitrary, however, in this paper we let $q_2 = 0.75$ and $q_3 = 0.25$, i.e. the third and the first quartile, respectively. This is motivated by the statistical properties of the model in which small fluctuations are driven by the base regime dynamics. Only the large deviations should be driven by the spike or drop regime dynamics.

The deseasonalized prices X_t and the conditional probabilities of being in the spike $P(R_t = 2 \mid x_1, x_2, \ldots, x_T)$ or drop $P(R_t = 3 \mid x_1, x_2, \ldots, x_T)$ regime for the analyzed datasets are displayed in Figs. 7 and 8. The prices classified as spikes or drops, i.e. with $P(R_t = 2 \mid x_1, x_2, \ldots, x_T) > 0.5$ or $P(R_t = 3 \mid x_1, x_2, \ldots, x_T) > 0.5$, are additionally denoted by dots or '×'. The estimated model parameters are given in Table 6.

The obtained base regime parameters are consistent with the well known properties of electricity prices. In particular, $\beta_1 \in [0.20, 0.44]$ indicates a relatively high speed of mean-reversion, while positive values of γ are responsible for the 'inverse leverage effect'. Finally, considering probabilities p_{ii} of staying in the same regime we obtain quite high values for each of the regimes, ranging from 0.6390 for the spike regime in the NSW market up to 0.9539 for the base regime in the same market. As

Table 7 Goodness-of-fit statistics for MRS models fitted to the deseasonalized EEX prices and NSW log-prices. For quantiles the relative differences between the sample and the model implied statistics are given (the latter were obtained from 100 simulations)

| | Quantiles | | | | | K–S test p-values | | | |
	0.1	0.25	0.5	0.75	0.9	Base	Spike	Drop	Model
EEX	−0.8%	−0.6%	−0.3%	0.2%	0.5%	0.50	0.26	0.69	0.31
NSW	−0.2%	−0.8%	−0.3%	0.8%	0.7%	0.09	0.33	–	0.18

a consequence, on average there are many consecutive observations from the same regime.

In order to check the statistical adequacy of the fitted MRS models we calculate percentage differences between the data and the model implied quantiles, see Table 7. The model implied values are obtained as the mean value of the statistics calculated over 100 simulated trajectories. A negative sign indicates that the value obtained from the dataset is lower than the model-implied. Observe that all differences between the data and the model-implied statistics are less than 0.8%, which indicates a relatively good fit of the models.

Moreover, we report the p-values of a Kolmogorov–Smirnov (K–S) goodness-of-fit type test for each of the individual regimes, as well as, for the whole model (for test details see Janczura and Weron). The goodness-of-fit results are summarized in Table 7. All K–S test p-values are higher than the commonly used 5% significance level, hence we cannot reject the hypotheses that the datasets follow the fitted MRS models.

6 Conclusions

In this paper we have proposed a method that greatly reduces the computational burden induced by the introduction of independent regimes in Markov regime-switching (MRS) models. Instead of storing conditional probabilities for each of the possible state process paths, our method requires conditional probabilities for only one time-step. This allows for a 100 to over 1000 times faster calibration than in case of a competing approach utilizing probabilities of the last 10 observations (see Table 5). Our method is also more general and admits any value of γ in the base regime dynamics. Since MRS models can be considered as generalizations of hidden Markov models (HMM), the presented results can have far-reaching implications for many problems where HMM have been applied (see e.g. Mamon and Elliott 2007). In some cases, perhaps, a MRS model with independent regimes would constitute a more realistic model of the observed phenomenon than a HMM.

We performed a limited simulation study to test the accuracy of the new method (introduced in Sect. 3) and applied it to sample series of electricity spot prices. The simulation study has shown that all sample means are close to the true parameter values (and all true parameter values are within the obtained 95% confidence intervals). Looking at the means, in most cases a sample of 1000 (or even 500 for the MR and IMR-G models; for model acronyms and definitions see Sect. 4) observations yields

satisfactory results, as the deviation does not exceed 0.03 (in absolute terms). Especially for the IMR-G model the results are very satisfactory. This is important in view of the fact that variants of this model are popular in the energy economics literature. In particular, a model of this type is calibrated in Sect. 5 to sample series of deseasonalized wholesale prices from the German EEX and Australian NSW markets. The model fits market data well and also replicates the major stylized facts of electricity spot price dynamics. In particular, the parameter γ can be treated as a parameter representing the 'degree of inverse leverage'. A positive value (e.g. 0.18 or 1.60 as in Table 6) indicates 'inverse leverage', which reflects the observation that positive electricity price shocks increase volatility more than negative shocks.

This paper does not resolve, however, all problems encountered when modeling wholesale electricity spot prices. In particular, two important issues should be addressed in future research: forecasting of the wavelet-based long-term seasonal component (LTSC) and recovery of the market observed term structure of volatility.

As mentioned in Sect. 5, predicting the wavelet-based LTSC beyond the next few weeks is a difficult task. This is because individual wavelet functions are localized in time and, unlike sines and cosines, decay to zero for longer time horizons. A much better in-sample description of the trend-seasonal behavior of electricity prices (which pretty well reflects the 'average' fuel price level understood as a combination of natural gas, crude oil and coal prices; see Janczura and Weron 2010) is achieved at the cost of a limitation in extrapolating the LTSC far into the future. On the other hand, sinusoidal or other periodic annual patterns can easily yield LTSC forecasts for the next few years … but are not very realistic for the stochastic and not too periodic behavior of the trend-seasonal component of electricity prices (as can be seen, for instance, in Fig. 6). What options are we left with then? One would be to use a periodic annual pattern for the LTSC and specify a generally much more complex spot price model to adequately describe the deviations of the market observed prices from the periodic LTSC. Another potentially promising approach would be to use a wavelet-based LTSC to obtain a relatively simple spot price model (as in this article) and use forward looking information, like smoothed electricity or fuel forward curves, to obtain LTSC forecasts beyond the next few weeks. However, this approach has its own limitations (data availability, estimation of the risk premium, etc.) and it has yet to be tested whether it provides a feasible forecasting tool.

The second important issue is the recovery of the market observed term structure of volatility, especially if the spot price models are to be used for derivatives valuation. In particular, the electricity forward prices implied by the spot price model should exhibit the so-called Samuelson effect—the decrease in volatility with increasing time to maturity—but the rate of decrease should be large only for maturities up to a year (Kiesel et al. 2009). Using Ito's lemma applied to $X_t e^{\beta t}$ and calculating the forward price as the expected future spot price we can calculate the volatility of forward prices within the MRS spot price models with mean-reverting base regime dynamics (2) and independent spikes and drops. It turns out that the volatility scales as $e^{-\beta(T-t)}$, analogously as in the one-factor model of Schwartz (1997), and consequently exhibits the Samuelson effect. However, the rate of decrease is completely determined by the speed of mean-reversion β. Moreover, exponential decay of the volatility and perfect correlation between forward prices (which is present in a single-

factor setting) is generally not consistent with observed electricity forward prices. Future research should address this limiting assumption. Perhaps, incorporating another stochastic factor would lead to a more realistic forward price curve.

Appendix

The recursive formula (12) can be derived in the following way. Let $\mathbf{X}_t = (X_1, X_2, \ldots, X_t)$. Observe that

$$X_{t,i} = \mathbb{I}_{R_t=i} X_t + \mathbb{I}_{R_t \neq i}\big[\alpha_i + (1 - \beta_i)X_{t-1,i} + \sigma_i |X_{t-1,i}|^{\gamma_i}\epsilon_t\big], \qquad (17)$$

where $\mathbb{I}_x$ is the indicator function. Taking the expected value conditional on $\mathbf{X}_t$ yields

$$
\begin{aligned}
E\big(X_{t,i} \mid \mathbf{X}_t; \theta^{(n)}\big) &= P\big(R_t = i \mid \mathbf{X}_t; \theta^{(n)}\big)X_t + P\big(R_t \neq i \mid \mathbf{X}_t; \theta^{(n)}\big)\big[\alpha_i^{(n)} \\
&\quad + \big(1 - \beta_i^{(n)}\big)E\big(X_{t-1,i} \mid \mathbf{X}_t, R_t \neq i; \theta^{(n)}\big) \\
&\quad + \sigma_i^{(n)} E\big(|X_{t-1,i}|^{\gamma_i^{(n)}}\epsilon_t \mid \mathbf{X}_t, R_t \neq i; \theta^{(n)}\big)\big].
\end{aligned}
$$

Since $X_{t-1,i}$ and ϵ_t are independent of the σ-algebra generated by $\{X_t, R_t \neq i\}$, we have

$$E\big(|X_{t-1,i}|^{\gamma_i^{(n)}}\epsilon_t \mid \mathbf{X}_t, R_t \neq i; \theta^{(n)}\big) = E\big(|X_{t-1,i}|^{\gamma_i^{(n)}}\epsilon_t \mid \mathbf{X}_{t-1}; \theta^{(n)}\big),$$

and

$$E\big(X_{t-1,i} \mid \mathbf{X}_t, R_t \neq i; \theta^{(n)}\big) = E\big(X_{t-1,i} \mid \mathbf{X}_{t-1}; \theta^{(n)}\big).$$

Moreover, from the law of iterated expectation and basic properties of conditional expected values:

$$
\begin{aligned}
&E\big(|X_{t-1,i}|^{\gamma_i^{(n)}}\epsilon_t \mid \mathbf{X}_{t-1}; \theta^{(n)}\big) \\
&= E\big[E\big(|X_{t-1,i}|^{\gamma_i^{(n)}}\epsilon_t \mid \mathbf{X}_{t-1}, X_{t-1,i}; \theta^{(n)}\big) \mid \mathbf{X}_{t-1}; \theta^{(n)}\big] \\
&= E\big[|X_{t-1,i}|^{\gamma_i^{(n)}} E\big(\epsilon_t \mid \mathbf{X}_{t-1}, X_{t-1,i}; \theta^{(n)}\big) \mid \mathbf{X}_{t-1}; \theta^{(n)}\big] \\
&= E\big[|X_{t-1,i}|^{\gamma_i^{(n)}} E(\epsilon_t) \mid \mathbf{X}_{t-1}; \theta^{(n)}\big] = 0,
\end{aligned}
$$

which implies

$$E\left(X_{t,i} \mid \mathbf{X}_t; \theta^{(n)}\right) = P\left(R_t = i \mid \mathbf{X}_t; \theta^{(n)}\right) X_t + P\left(R_t \neq i \mid \mathbf{X}_t; \theta^{(n)}\right)$$
$$\cdot \left[\alpha_i^{(n)} + \left(1 - \beta_i^{(n)}\right) E\left(X_{t-1,j} \mid \mathbf{X}_{t-1}; \theta^{(n)}\right)\right].$$

Finally, substituting the variables $\mathbf{X}_t$ with their observations $\mathbf{x}_t$ leads to formula (12).

References

Benth, F.E., Koekebakker, S., Ollmar, F.: Extracting and applying smooth forward curves from average-based commodity contracts with seasonal variation. J. Deriv. 52–66 (2007)

Benth, F.E., Benth, J.S., Koekebakker, S.: Stochastic Modeling of Electricity and Related Markets. World Scientific, Singapore (2008a)

Benth, F.E., Cartea, A., Kiesel, R.: Pricing forward contracts in power markets by the certainty equivalence principle: explaining the sign of the market risk premium. J. Bank. Finance 32(10), 2006–2021 (2008b)

Bierbrauer, M., Menn, C., Rachev, S.T., Trück, S.: Spot and derivative pricing in the EEX power market. J. Bank. Finance 31, 3462–3485 (2007)

Borak, S., Weron, R.: A semiparametric factor model for electricity forward curve dynamics. J. Energy Mark. 1(3), 3–16 (2008)

Cappe, O., Moulines, E., Ryden, T.: Inference in Hidden Markov Models. Springer, Berlin (2005)

Christensen, T., Hurn, S., Lindsay, K.: It never rains but it pours: modeling the persistence of spikes in electricity prices. Energy J. 30(1), 25–48 (2009)

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

De Jong, C.: The nature of power spikes: a regime-switch approach. Stud. Nonlinear Dyn. Econom. 10(3), 3 (2006)

Dempster, A., Laird, N., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. 39, 1–38 (1977)

Erlwein, C., Benth, F.E., Mamon, R.: HMM filtering and parameter estimation of an electricity spot price model. Energy Econ. 32, 1034–1043 (2010)

Eydeland, A., Wolyniec, K.: Energy and Power Risk Management. Wiley, Hoboken (2003)

Fink, G.A.: Markov Models for Pattern Recognition: From Theory to Applications. Springer, Berlin (2008)

Gray, S.F.: Modeling the conditional distribution of interest rates as a regime-switching process. J. Financ. Econ. 42, 27–62 (1996)

Hahn, M., Frühwirth-Schnatter, S., Sass, J.: Estimating models based on Markov jump processes given fragmented observation series. AStA Adv. Stat. Anal. 93, 403–425 (2009)

Hamilton, J.: A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57, 357–384 (1989)

Hamilton, J.: Analysis of time series subject to changes in regime. J. Econom. 45, 39–70 (1990)

Hamilton, J.: Regime switching models. In: The New Palgrave Dictionary of Economics, 2nd edn. (2008)

Huisman, R.: An Introduction to Models for the Energy Markets. Risk Books, London (2009)

Huisman, R., de Jong, C.: Option formulas for mean-reverting power prices with spikes. ERIM Report Series Reference No. ERS-2002-96-F&A (2002)

Huisman, R., Mahieu, R.: Regime jumps in electricity prices. Energy Econ. 25, 425–434 (2003)

Janczura, J., Weron, R.: An empirical comparison of alternate regime-switching models for electricity spot prices. Energy Econ. 32, 1059–1073 (2010)

Janczura, J., Weron, R.: Goodness-of-fit testing for the marginal distribution of regime-switching models. Available at MPRA: http://mpra.ub.uni-muenchen.de/32532 (2011)

Kanamura, T., Ōhashi, K.: On transition probabilities of regime switching in electricity prices. Energy Econ. 30, 1158–1172 (2008)

Karakatsani, N.V., Bunn, D.W.: Intra-day and regime-switching dynamics in electricity price formation. Energy Econ. 30, 1776–1797 (2008)

Karakatsani, N.V., Bunn, D.: Fundamental and behavioural drivers of electricity price volatility. Stud. Nonlinear Dyn. Econom. 14(4), 4 (2010)

Kholodnyi, V.A.: Modeling power forward prices for power with spikes: a non-Markovian approach. Nonlinear Anal. 63, 958–965 (2005)

Kiesel, R., Schindlmayr, G., Börger, R.H.: A two-factor model for the electricity forward market. Quant. Finance **9**(3), 279–287 (2009)

Kim, C.-J.: Dynamic linear models with Markov-switching. J. Econom. **60**, 1–22 (1994)

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

Luo, Q., Mao, X.: Stochastic population dynamics under regime switching. J. Math. Anal. Appl. **334**(1), 69–84 (2007)

Mamon, R.S., Elliott, R.J. (eds.): Hidden Markov models in finance. In: International Series in Operations Research & Management Science, vol. 104. Springer, Berlin (2007)

Mari, C.: Random movements of power prices in competitive markets: a hybrid model approach. J. Energy Mark. **1**(2), 87–103 (2008)

Mount, T.D., Ning, Y., Cai, X.: Predicting price spikes in electricity markets using a regime-switching model with time-varying parameters. Energy Econ. **28**, 62–80 (2006)

Schwartz, E.S.: The stochastic behavior of commodity prices: implications for valuation and hedging. J. Finance **52**(3), 923–973 (1997)

Trück, S., Weron, R., Wolff, R.: Outlier treatment and robust approaches for modeling electricity spot prices. In: Proceedings of the 56th Session of the ISI (2007). Available at MPRA: http://mpra.ub.uni-muenchen.de/4711/

Weron, R.: Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach. Wiley, Chichester (2006)

Weron, R.: Market price of risk implied by Asian-style electricity options and futures. Energy Econ. **30**, 1098–1115 (2008)

Weron, R.: Heavy-tails and regime-switching in electricity prices. Math. Methods Oper. Res. **69**(3), 457–473 (2009)

Wong, H., Ip, W.-C., Xie, Z., Lui, X.: Modelling and forecasting by wavelets, and the application to exchange rates. J. Appl. Stat. **30**(5), 537–553 (2003)

Yousefi, S., Weinreich, I., Reinarz, D.: Wavelet-based prediction of oil prices. Chaos Solitons Fractals **25**, 265–275 (2005)