HSC Research Report

HSC/12/01

Inference for Markov-regime switching models of electricity spot prices

Joanna Janczura*
Rafał Weron**

* Hugo Steinhaus Center, Wrocław University of Technology, Poland
** Institute of Organization and Management, Wrocław University of Technology, Poland

Hugo Steinhaus Center
Wrocław University of Technology
Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland
http://www.im.pwr.wroc.pl/~hugo/

Inference for Markov-regime switching models of electricity spot prices

Joanna Janczura and Rafał Weron

1 Introduction

The basic idea that underlies Markov regime-switching (MRS) is that of representing the behavior of an observed time series by separate states or regimes, which can be driven by different stochastic processes. Unlike threshold type regime-switching models (e.g. TAR, STAR, SETAR), in MRS models the regimes are only latent and, hence, these models do not require an upfront specification of the threshold variable and level. This flexible specification has led to their popularity not only in econometrics (Choi, 2009; Hamilton, 2008) but also in other fields of science including traffic modeling (Cetin and Comert, 2006), population dynamics (Luo and Mao, 2007), river flow analysis (Vasas et al., 2007) and pattern recognition (Fink, 2008).

In energy economics MRS models have seen extensive use due to their ability to capture the unique behavior of electricity spot prices (Bierbrauer et al., 2004, 2007; De Jong, 2006; Erlwein et al., 2010; Hirsch, 2009; Huisman, 2009; Huisman and Mahieu, 2003; Janczura and Weron, 2010, 2012a; Kanamura and Ohashi, 2008; Karakatsani and Bunn, 2008; Kholodnyi, 2005; Mari, 2008; Mount et al., 2006; Weron, 2009). And recall that electricity is a very specific commodity. Firstly, it is practically non-storable and requires immediate delivery. Secondly, end-user demand is weather and business cycle dependent. Thirdly, effects like power plant outages, transmission grid reliability and strategic bidding add complexity and randomness. The resulting spot prices exhibit significant seasonality on the annual, weekly and daily level, as well as, mean reversion, very high volatility and generally short-lived extreme price spikes and/or drops.

Joanna Janczura
Hugo Steinhaus Center, Institute of Mathematics and Computer Science, Wrocław University of Technology, 50-370 Wrocław, Poland, e-mail: joanna.janczura@pwr.wroc.pl

Rafał Weron (*corresponding author*)
Institute of Organization and Management, Wrocław University of Technology, 50-370 Wrocław, Poland, e-mail: rafal.weron@pwr.wroc.pl

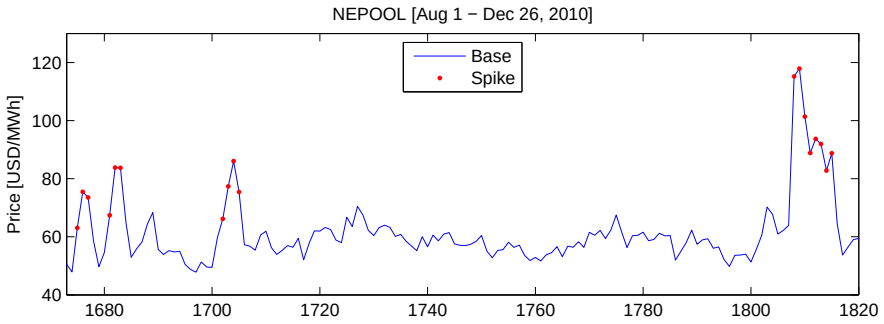

Fig. 1 Deseasonalized mean daily electricity spot price from the New England power market (NEPOOL, U.S.) in the period August 1 – December 26, 2010. The prices classified as spikes are denoted by dots (see Section 5 for deseasonalization and model details). The regime switches and spike clustering are clearly visible.

These extreme price movements tend to cluster (Bierbrauer et al., 2004; Christensen et al., 2009; Janczura and Weron, 2010, see also Figure 1), which makes the very popular class of jump-diffusion models impractical, as they cannot exhibit consecutive spikes with the frequency observed in power market data (Weron et al., 2004). On the other hand, 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 class of 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, in the energy economics literature the base regime is typically modeled by a mean-reverting diffusion (Benth et al., 2008; Huisman, 2009; Weron, 2006), sometimes heteroskedastic (Janczura and Weron, 2009), while for the spike (or drop) regime(s) a number of specifications have been suggested, ranging from mean-reverting diffusions to heavy tailed random variables (for a review see Janczura and Weron, 2010).

After selecting 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 – hence the alternative name 'parameter-switching'; an approach dating back to 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 better choice for electricity spot price processes, which can exhibit a moderately volatile and symmetric (in terms of the marginal distribution) behavior in the base regime and a very volatile and an asymmetric one in the spike regime, see Figure 1. We will look more closely at these independent regime models in Section 2.

Once the electricity spot price model is specified we are left with the problem of calibrating it to market data. Due to the unobservable switching mechanism, esti-

mation 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. We have recently proposed a method that greatly reduces the computational burden in such a case (Janczura and Weron, 2012a). As we will see in Section 3, the method allows for a 100 to over 1000 times faster calibration than a competing approach utilizing probabilities of the last 10 observations. Instead of storing conditional probabilities for each of the possible state process paths, it 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 also for many problems where HMM have been applied (see e.g. Mamon and Elliott, 2007). In Section 3 we will also show that the fit can be further improved by optimizing the cutoff(s) used for separating the regimes, instead of arbitrarily setting them to the median (Janczura and Weron, 2010) or the 1st and 3rd quartiles (Janczura and Weron, 2012a) of the deseasonalized dataset.

While the existence of distinct regimes in electricity prices is generally unquestionable (being a consequence of the non-linear, heterogeneous supply stack structure in the power markets, see e.g. Eydeland and Wolyniec, 2012; Weron, 2006), the actual goodness-of-fit of the models requires statistical validation. However, recent work concerning the statistical fit of regime-switching models has been mainly devoted to testing parameter stability versus the regime-switching hypothesis. Several tests have been constructed for the verification of the number of regimes. Most of them exploit the likelihood ratio technique (Cho and White, 2007; Garcia, 1998), but there are also approaches related to recurrence times (Sen and Hsieh, 2009), likelihood criteria (Celeux and Durand, 2008) or the information matrix (Hu and Shin, 2008). Specification tests, like tests for omitted autocorrelation or omitted explanatory variables based on the score function technique, were proposed earlier by Hamilton (1996). On the other hand, procedures for goodness-of-fit testing of the marginal distribution of regime-switching models have been derived only recently. Janczura and Weron (2012b) have proposed two empirical distribution function (edf) based testing techniques built on the Kolmogorov-Smirnov test. As we will see in Section 4, the procedure is readily applicable to regime-switching models of electricity spot prices.

We conclude this paper with applications of the presented techniques to wholesale electricity prices from two major power markets – the German EEX and the North American NEPOOL (Section 5). Finally, in the Conclusions we summarize the presented results and provide suggestions for future work in this interesting area.

2 Regime-switching models

Assume that the observed process X_t may be in one of L states (regimes) at time t, dependent on the state process R_t:

$$X_t = \begin{cases} X_{t,1} & \text{if } R_t = 1, \\ X_{t,2} & \text{if } R_t = 2, \\ \vdots & \vdots \quad \vdots \\ X_{t,L} & \text{if } R_t = L. \end{cases} \tag{1}$$

Possible specifications of the process R_t may be divided into two classes: those where the current state of the process is observable (like threshold models, e.g. TAR, SETAR) and those where it is latent. Probably the most prominent representatives of the second group are the hidden Markov models (HMM; for a review see e.g. Cappe et al., 2005) and their generalizations allowing for temporal dependence within the regimes – the Markov regime-switching models (MRS). Like in HMM, in MRS models R_t is assumed to be a Markov chain governed by the transition matrix $\mathbf{P}$ containing the probabilities p_{ij} of switching from regime i at time t to regime j at time $t+1$, for $i, j = \{1, 2, ..., L\}$:

$$\mathbf{P} = (p_{ij}) = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1L} \\ p_{21} & p_{22} & \cdots & p_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ p_{L1} & p_{L2} & \cdots & p_{LL} \end{pmatrix}, \text{ with } p_{ii} = 1 - \sum_{j \neq i} p_{ij}. \tag{2}$$

The current state R_t at time t depends on the past only through the most recent value R_{t-1}. Consequently, the probability of being in regime j at time $t+m$ starting from regime i at time t is given by

$$P(R_{t+m} = j \mid R_t = i) = (\mathbf{P}')^m \cdot e_i, \tag{3}$$

where $\mathbf{P}'$ denotes the transpose of $\mathbf{P}$ and e_i is the ith column of the identity matrix.

The definitions of the individual regimes can be arbitrarily chosen depending on the modeling needs. In this paper we focus on the independent regime (spike) model (Bierbrauer et al., 2007; De Jong, 2006; Huisman and de Jong, 2003; Janczura and Weron, 2010), as it seems to be a reasonable choice for electricity spot price processes which can exhibit qualitatively different dynamics in each regime. At the same time, however, it is more computationally challenging than the popular parameter-switching model (for a detailed description of the latter we refer to Janczura and Weron, 2012a).

In the independent regime (spike) model, X_t is defined by (1) with at least one regime given by:

$$X_{t,i} = \alpha_i + (1 - \beta_i) X_{t-1,i} + \sigma_i |X_{t-1,i}|^{\gamma_i} \varepsilon_{t,i}, \tag{4}$$

where $\alpha_i, \beta_i, \sigma_i$ and γ_i are constants and $\varepsilon_{t,i}$'s are i.i.d. Gaussian random variables. The absolute value in the above formula is needed if negative data is analyzed. Note, that formula (4) is a discrete-time version of the mean-reverting, possibly heteroskedastic process given by the following Ornstein-Uhlenbeck-type stochastic differential equation:

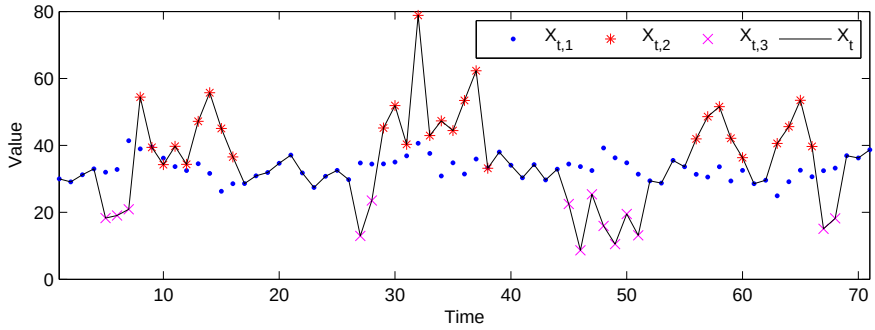

Fig. 2 A sample trajectory of a MRS model with three independent regimes (black solid line) superimposed on the observable and latent values of the processes in the regimes. Observe, that the values $X_{t,1}$ of the mean-reverting regime become latent when the process is in another state. The simulation was performed for a 3-regime model defined by eqns. (24)-(27), see Section 5, with the following parameters: $\mathbf{P} = (p_{ij}) = [0.90, 0.05, 0.05;\ 0.25, 0.70, 0.05;\ 0.25, 0.05, 0.70]$, $\alpha_1 = 10$, $\beta_1 = 0.3$, $\sigma_1^2 = 20$, $\gamma_1 = 0$, $\alpha_2 = 2.5$, $\sigma_2^2 = 0.5$, $\alpha_3 = 2.5$, $\sigma_3^2 = 0.5$, $q_2 = q_3 = 30$.

$$dX_t = (\alpha - \beta X_t)dt + \sigma |X_t|^{\gamma} dW_t, \tag{5}$$

where W_t is the Wiener process. The remaining regimes constitute i.i.d. samples from specified continuous, strictly monotone distributions F^i:

$$X_{t,i} \sim F^i(x). \tag{6}$$

An example of such a specification is the 3-regime model with mean-reverting, heteroskedastic base regime (i.e. 'normal' prices) dynamics and independent spikes and drops, as proposed by Janczura and Weron (2010). In Section 5 we apply it to electricity spot prices from the EEX and NEPOOL power markets. Note, that in such a model the values of the mean reverting regime become latent when the process is in another state, see Figure 2 for an illustration.

3 Calibration

Calibration of MRS models is not straightforward since the regimes are 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 Section 3.1 we briefly describe the general estimation procedure. Next, in Section 3.2 we discuss the computational problems induced by the introduction of independent regimes and present an efficient remedy. Finally, in Section 3.4 we show that the fit can be further improved by optimizing the cut-off(s) used for separating the regimes.

3.1 Expectation-Maximization (EM) algorithm

The algorithm starts with an arbitrarily chosen vector of initial parameters $\theta^{(0)} = (\eta^{(0)}, \mathbf{P}^{(0)}, \rho_i^{(0)})$, for $i = 1,2,...,L$, where $\rho_i^{(0)} \equiv P(R_1 = i)$ and $\eta^{(0)}$ is a vector of parameters defined by equations (4) and (6). 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 $E(\mathbb{I}_{R_t=i}|x_1, x_2..., x_T; \theta)$, can be calculated. These expectations result in the so called 'smoothed inferences', i.e. the conditional probabilities $P(R_t = j|x_1, ..., x_T; \theta)$ 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 $\mathbf{x}_t = (x_1, x_2, ...x_t)$. The E-part consists of the following steps (Kim, 1994):

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

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

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

and

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

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

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

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

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

3.1.2 The M-step

In the second step of the EM algorithm, new and more exact maximum likelihood (ML) estimates $\eta^{(n+1)}$ for all model parameters are calculated. Compared to standard ML estimation, where for a given pdf f the log-likelihood function $\sum_{t=1}^{T} \log f(x_t, \eta^{(n+1)})$ 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 $P(R_t = i|\mathbf{x}_T; \theta^{(n)})$. Namely, the ML estimates are derived maximizing the log-likelihood function of the following form:

$$\log\left[L(\eta^{(n+1)})\right] = \sum_{i=1}^{L}\sum_{t=1}^{T} P(R_t = i|\mathbf{x_T}; \theta^{(n)}) \log\left[f(x_t|R_t = i, \mathbf{x_{t-1}}; \eta^{(n+1)})\right]. \quad (7)$$

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

$$\begin{aligned} p_{ij}^{(n+1)} &= \frac{\sum_{t=2}^{T} P(R_t = j, R_{t-1} = i|\mathbf{x}_T; \theta^{(n)})}{\sum_{t=2}^{T} P(R_{t-1} = i|\mathbf{x}_T; \theta^{(n)})} = \\ &= \frac{\sum_{t=2}^{T} P(R_t = j|\mathbf{x}_T; \theta^{(n)}) \frac{p_{ij}^{(n)} P(R_{t-1}=i|\mathbf{x}_{t-1};\theta^{(n)})}{P(R_t=j|\mathbf{x}_{t-1};\theta^{(n)})}}{\sum_{t=2}^{T} P(R_{t-1} = i|\mathbf{x}_T; \theta^{(n)})}, \end{aligned} \quad (8)$$

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 $\theta^{(n+1)} = (\eta^{(n+1)}, \mathbf{P}^{(n+1)}, \rho_i^{(n+1)})$, $i = 1, 2, ..., L$, in the next iteration of the E-step.

3.2 Independent regimes

Both steps of the EM algorithm require derivation of the conditional probability density functions $f(x_t|R_t = i; \mathbf{x}_{t-1}; \theta^{(n)})$. For the regime(s) described by i.i.d. random variables, see eqn. (6), this is just the model specified pdf. However, for the mean-reverting regime(s), see eqn. (4), the situation is more complicated due to the dependence structure of the driving process. If the regimes are independent from each other, the values of the mean-reverting regime become latent when the process is in the other states, see Figure 2. This makes the distribution of X_t dependent on the whole history $(x_1, x_2, ..., x_{t-1})$ of the process. As a consequence, all possible paths of the state process $(R_1, R_2, ..., R_t)$ should be considered in the derivation of the pdf, implying that $f(x_t|R_t = i, R_{t-1} \neq i, ..., R_{t-j} \neq i, R_{t-j-1} = i; \mathbf{x}_{t-1}; \theta^{(n)})$ and the whole

set of probabilities $P(R_t = i_t, R_{t-1} = i_{t-1}, ..., R_{t-j} = i_{t-j} | \mathbf{x}_{t-1}; \theta^{(n)})$ should be used in the EM algorithm.

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 other regimes is negligible.

Instead, following Janczura and Weron (2012a), we suggest to approximate the latent variables $x_{t-1,i}$ from the mean-reverting regimes by their expectations $\tilde{x}_{t-1,i} = E(X_{t-1,i}|\mathbf{x}_{t-1}; \theta^{(n)})$ 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. Note that if $x_{t-1,i}$ was observable, then X_t given $R_t = i$ and $x_{t-1,i}$ would be Gaussian distributed with mean $(1-\beta_i^{(n)})x_{t-1,i} + \alpha_i^{(n)}$ and variance $(\sigma_i^{(n)})^2 |x_{t-1,i}|^{2\gamma_i^{(n)}}$. Hence, the estimation procedure described in Section 3.1 can be applied with the following approximation of the mean-reverting regime pdf:

$$f\left(x_t | R_t = i; \mathbf{x}_{t-1}; \theta^{(n)}\right) = \frac{1}{\sqrt{2\pi}\sigma_i^{(n)} |\tilde{x}_{t-1,i}|^{\gamma_i^{(n)}}} \cdot \cdot \exp\left\{ - \frac{\left(x_t - \left(1-\beta_i^{(n)}\right)\tilde{x}_{t-1,i} - \alpha_i^{(n)}\right)^2}{2\left(\sigma_i^{(n)}\right)^2 |\tilde{x}_{t-1,i}|^{2\gamma_i^{(n)}}} \right\}. \tag{9}$$

The expected values $\tilde{x}_{t,i} = E(X_{t,i}|\mathbf{x}_t; \theta^{(n)})$ can be computed using the following recursive formula (for the derivation see Janczura and Weron, 2012a):

$$E\left(X_{t,i}|\mathbf{x}_t; \theta^{(n)}\right) = P\left(R_t = i|\mathbf{x}_t; \theta^{(n)}\right) x_t + P\left(R_t \neq i|\mathbf{x}_t; \theta^{(n)}\right) \cdot \cdot \left\{ \alpha_i^{(n)} + \left(1-\beta_i^{(n)}\right) E\left(X_{t-1,i}|\mathbf{x}_{t-1}; \theta^{(n)}\right) \right\}. \tag{10}$$

Moreover, these expected values are linear combinations of the observed vector $\mathbf{x}_t$ and the probabilities $P(R_j = i|\mathbf{x}_j; \theta^{(n)})$ calculated during the estimation procedure (see the filtering part of the E-step):

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

Hence, by using $\tilde{x}_{t-1,i} = E(X_{t-1,i}|\mathbf{x}_{t-1};\theta^{(n)})$ in formula (9), instead of x_{t-1}, the computational complexity of the E-step is greatly reduced. In fact, the total number of probabilities stored in computer memory is now 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.

3.3 Time-varying transition probabilities

The independent regime models discussed above can provide adequate fits to electricity spot prices in terms of the marginal distributions, but not in terms of the temporal behavior. As Mount et al. (2006) and Cartea et al. (2009) have shown, the timing of spikes could be improved by incorporating forward looking information on capacity constraints. Unfortunately, the availability (to every market participant) of the reserve margin data is limited. If temperature is used as a proxy for the reserve margin (as in Huisman, 2008), the results are not as good.

A relatively simple, yet potentially rewarding alternative is to admit a transition matrix with time-varying probabilities of a one year period: $p_{ij}(t) = p_{ij}(t+1\text{year})$. Following Janczura and Weron (2010) the probabilities can be calibrated in a two-step procedure in the last part of the E-step of the EM algorithm. First, the probabilities are estimated independently for each of the four seasons: Winter (months XII-II), Spring (III-V), Summer (VI-VIII) and Autumn (IX-XI). Then they are smoothed using a kernel density estimator with a Gaussian kernel. More complex annual structures and smoothing techniques can be used as well. Here, however, for simplicity we will limit the analysis to the original approach.

3.4 Optimizing the cutoffs

To eliminate spike misclassification in some early MRS models, including the unwanted feature of negative 'expected spike sizes', i.e. $E(X_{t,spike}) < E(X_{t,base})$, Janczura and Weron (2009) proposed to use median-shifted spike regime distributions. This was motivated by a common-sense assumption that small fluctuations

should be driven by the base regime dynamics and only the large deviations by the spike (or drop) regime dynamics. For EEX market data from the period 2001-2009 they found that models with shifted spike regime distributions (which assign zero probability to prices below the median of deseasonalized prices) led to more realistic descriptions of electricity spot prices.

Originally, Janczura and Weron (2009) introduced median-shifted log-normal:

$$\log(X_{t,i} - X(q_i)) \sim \mathrm{N}(\alpha_i, \sigma_i^2), \quad X_{t,i} > X(q_i), \tag{11}$$

and Pareto:

$$X_{t,i} \sim F_{\text{Pareto}}(\sigma_i, \alpha_i) = 1 - \left(\frac{\alpha_i}{x}\right)^{\sigma_i}, \quad x > \alpha_i \geq X(q_i), \tag{12}$$

spike regime distributions, but the latter was found to be too heavy-tailed for the analyzed datasets (Janczura and Weron, 2010). 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, for simplicity it can be set to the median (which can be interpreted as a value representing the average capacity margin in a power market; when the price exceeds this value the spikes occur) or a quartile (e.g. 1st for the drop and 3rd for the spike regime, as in Janczura and Weron, 2012a) of the deseasonalized dataset.

Nothing, however, prevents us from optimizing these cutoff levels, both for the spike and drop regimes. For the 3-regime model studied in Section 5, this can be achieved by running a 2-dimensional optimization (e.g. using the Nelder-Mead simplex routine in Matlab) with the objective of maximizing the likelihood. Precisely, for given cutoff levels the MRS model is calibrated and the log-likelihood function is evaluated. Next, the log-likelihood is treated as a function of the cutoffs and the optimization procedure is performed.

The computational cost is not overwhelming – typically under 100 calibrations of the MRS model have to be performed before a (local) maximum is reached, using the default parameters of the simplex routine in Matlab. Increasing the termination tolerance can naturally greatly speed up the process, even without a significant loss of precision. In Section 5 we will check how well this optimization works and how different form the median or the quartiles are the obtained optimal cutoff levels.

4 Goodness-of-fit testing

The adequacy of the models can be evaluated on the base of descriptive statistics, as well as, goodness-of-fit hypothesis tests. The former include the Inter-Quartile and the Inter-Decile Range, i.e. the difference between the third and the first quartiles (IQR) or ninth and first deciles (IDR). The quantile-based measures rather than the less robust to outliers moment-related statistics are preferred (Janczura and Weron, 2010). A more sound decision can be made based on a goodness-of-fit test, tailored to evaluate the fit of regime-switching models. Here we briefly summarize the

methods proposed by Janczura and Weron (2012b); for derivations and performance evaluation we refer to the original paper. The methods are based on the Kolmogorov-Smirnov (K-S) goodness-of-fit test and verify whether the null hypothesis H_0 that observations come from the distribution implied by the model specification cannot be rejected. The procedure can be easily adapted to other empirical distribution function (edf) type tests, like the Anderson-Darling test (see e.g. D'Agostino and Stevens, 1986). For clarity of exposition we limit the discussion in this Section to 2-regime models only with the first regime driven by a mean-reverting process and the second by an i.i.d. F^2-distributed sample. However, all presented results are also valid for $L > 2$.

Recall that the Kolmogorov-Smirnov test statistic is given by:

$$D_n = \sqrt{n} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|, \tag{13}$$

where n is the sample size, F_n is the empirical distribution function (edf) and F is the corresponding theoretical cumulative distribution function (cdf). Hence, having an i.i.d. sample $(y_1, y_2, \ldots, y_n)$, the test statistic can be calculated as

$$d_n = \sqrt{n} \max_{1 \le t \le n} \left| \sum_{k=1}^{n} \frac{1}{n} \mathbb{I}_{\{y_k \le y_t\}} - F(y_t) \right|, \tag{14}$$

where $\mathbb{I}$ is the indicator function. If hypothesis H_0 is true, then the statistic D_n asymptotically has the Kolmogorov-Smirnov distribution (KS). Therefore if n is large enough, the following approximation holds

$$P(D_n \ge c | H_0) \approx P(\kappa \ge c), \tag{15}$$

where $\kappa \sim KS$ and c is the critical value. Hence, the p-value for an i.i.d. sample $(y_1, y_2, \ldots, y_n)$ can be approximated by $P(\kappa \ge d_n)$.

4.1 The ewedf approach

The described above testing scheme is valid for i.i.d. samples. In order to apply it in the framework of MRS models, we have to overcome two problems. First, the regimes are only latent, so we cannot unambiguously distinguish observations from different regimes (and consequently from different distributions). Second, there is a dependence structure within the mean-reverting regime.

The first issue can be resolved by performing an identification of the state process. Recall that, as a result of the estimation procedure described in Section 3, the so called 'smoothed inferences' about the state process are derived. The smoothed inferences are the probabilities that the t-th observation comes from a certain regime given the whole available information $P(R_t = i | x_1, x_2, \ldots, x_T)$. Hence, a natural choice is to relate each observation with the most probable regime by

letting $R_t = i$ if $P(R_t = i|x_1, x_2, ..., x_T) > 0.5$. However, we have to mention, that the hypothesis H_0 now states that $(x_1, x_2, ..., x_T)$ is driven by a regime-switching model with known state process values. We call this approach 'ewedf', which stands for 'equally-weighted empirical distribution function' (Janczura and Weron, 2009, 2012b).

Now, we focus on the problem of dependence between mean-reverting regime observations. Provided that the values of the state process R_t are known, observations can be split into separate subsamples related to each of the regimes. Namely, subsample i consists of all values X_t satisfying $R_t = i$. The regimes are independent from each other, but the i.i.d. condition must be satisfied within the subsamples themselves. Therefore the mean-reverting regime observations are substituted by their respective residuals. Using the Euler scheme and rearranging terms of formula (5), we get that

$$\varepsilon_{t,1} = \frac{X_t - (1 - \beta_1 \Delta t) X_{t-\Delta t} - \alpha_1 \Delta t}{\sqrt{\Delta t} \sigma_1 |X_{t-\Delta t}|^{\gamma_1}}, \tag{16}$$

has the standard Gaussian distribution, where Δt is the time interval between consecutive mean-reverting regime observations. However, since the Euler scheme is an approximation of a continuous process, formula (16) is valid only for small Δt. In contrast, if the mean-reverting regime dynamics is given by the AR(1) process, i.e. the process defined by (4) with $\gamma = 0$, exact residuals can be derived. Precisely, the residuals are derived from all pairs of consecutive AR(1) observations as:

$$\varepsilon_{t,1} = \frac{x_t - (1 - \beta_1)^k x_{t-1} - \alpha_1 \frac{1-(1-\beta_1)^k}{\beta_1}}{\sigma_1 \sqrt{\frac{1-(1-\beta_1)^{2k}}{1-(1-\beta_1)^2}}}, \tag{17}$$

where $(k-1)$ is the number of latent observations from the mean reverting regime (or equivalently the number of observations from the second regime that occurred between two consecutive AR(1) observations) and α_1, β_1 and σ_1 are the model parameters, see (4).

Transformation (16), or (17) in the AR(1) case, ensures that the subsample containing observations from the mean-reverting regime is i.i.d. Since the second regime is i.i.d. by definition, the standard Kolmogorov-Smirnov test can be applied to each of the subsamples.

The goodness-of-fit of the marginal distribution of the individual regimes can be formally tested, using the test statistic (14). For the mean-reverting regime, F is the standard Gaussian cdf and $(y_1, y_2, ..., y_{n_1})$ is the subsample of the standardized residuals obtained by applying transformation (17), while for the other regimes, F is the model specified cdf (i.e. F^2) and $(y_1, y_2, ..., y_{n_2})$ is the subsample of respective observations. Observe that the 'whole model' goodness-of-fit can be also verified, using the fact that for $X \sim F^2$ we have that $Y = (F)^{-1}[F^2(X)]$ is F-distributed. Indeed, a sample $(y_1^1, y_2^1, ..., y_{n_1}^1, y_1^2, y_2^2, ..., y_{n_2}^2)$, where y_t^1's are the standardized residuals of the mean-reverting regime, while y_t^2's are the transformed variables corresponding

to the second regime, i.e. $y_t^2 = (F)^{-1}[F^2(x_{t,2})]$ with F being the standard Gaussian cdf, is i.i.d. $N(0,1)$-distributed and, hence, the testing procedure is applicable.

4.2 The wedf approach

Now, we briefly mention another potentially useful testing approach dealing with the latency of the state process. Observe, that in the standard goodness-of-fit testing approach based on the edf each observation is taken into account with weight $\frac{1}{n}$ (i.e. inversely proportional to the size of the sample). However, in MRS models the state process is latent. The estimation procedure (the EM algorithm) only yields the probabilities that a certain observation comes from a given regime. Moreover, in the resulting marginal distribution of the MRS model each observation is, in fact, weighted with the corresponding probability. Therefore, a similar approach could be used in the testing procedure. As Janczura and Weron (2012b) have shown, this is possible for independent regime models with homoskedastic mean-reverting dynamics, i.e. with $\gamma = 0$ in formula (4). The approach uses the concept of the weighted empirical distribution function (wedf):

$$F_n^w(x) = \sum_{t=1}^{n} \frac{w_t \mathbb{I}_{\{y_t < x\}}}{\sum_{t=1}^{n} w_t}, \tag{18}$$

where $(y_1, y_2, ..., y_n)$ is a sample of observations and $(w_1, ..., w_n)$ are the corresponding weights, such that $0 \leq w_t \leq M$, $\forall_{t=1,...,n}$. A natural choice of weights seems to be $w_t = P(R_t = i|x_1, x_2, ..., x_T) = E(\mathbb{I}_{\{R_t=i\}}|x_1, x_2, ..., x_T)$ for the i-th regime observations. Indeed, it can be shown that, if the H_0 hypothesis is true, the test statistic

$$D_n^w = \sqrt{n} \sup_{x \in \mathbb{R}} |F_n^w(x) - F(x)|, \tag{19}$$

converges weakly to the Kolmogorov-Smirnov distribution, with F_n^w derived for the sample $(y_1^1, y_2^1, ..., y_{T-1}^1, y_1^2, y_2^2, ..., y_T^2)$, where $(y_1^1, y_2^1, ..., y_{T-1}^1)$ are the transformed variables of the mean-reverting regime and $(y_1^2, y_2^2, ..., y_T^2)$ are the variables corresponding to the second regime, i.e. $y_t^2 = (F)^{-1}[F^2(x_t)]$ with F being the standard Gaussian cdf. The transformation of the mean-reverting regime observations is, similarly as in the ewedf approach, based on deriving the process residuals. We have:

$$\varepsilon_{t,1} = \frac{X_{t,1} - \alpha - (1-\beta)E(X_{t-1,1}|\mathbf{x}_{t-1})}{\sqrt{(1-\beta)^2 Var(X_{t-1,1}|\mathbf{x}_{t-1}) + \sigma^2}}, \tag{20}$$

where $E(X_{t-1,1}|\mathbf{x}_{t-1})$ and $Var(X_{t-1,1}|\mathbf{x}_{t-1})$ can be calculated using the following formulas:

$$\begin{aligned} E(X_{t,1}|\mathbf{x}_t) &= P(R_t = 1|\mathbf{x}_t)x_t + \\ &\quad + P(R_t \neq 1|\mathbf{x}_t)\left[\alpha + (1-\beta)E(x_{t-1,1}|\mathbf{x}_{t-1})\right], \end{aligned} \tag{21}$$

$$\begin{aligned} E(X_{t,1}^2|\mathbf{x}_t) &= P(R_t = 1|\mathbf{x}_t)x_t^2 + \\ &\quad + P(R_t \neq 1|\mathbf{x}_t)\big[\alpha^2 + 2\alpha(1-\beta)E(X_{t-1,1}|\mathbf{x}_{t-1}) + \\ &\quad + (1-\beta)^2 E(X_{t-1,1}^2|\mathbf{x}_{t-1}) + \sigma^2\big]. \end{aligned} \tag{22}$$

Finally, the p-value for the sample $(y_1^1, y_2^1, ..., y_{T-1}^1, y_1^2, y_2^2, ..., y_T^2)$ can be approximated by $P(\kappa \geq d_n)$, where

$$d_n = \sqrt{n} \max_{1 \leq t \leq n} \max_{i=1,2} \left| F_n^w(y_t^i) - F(y_t^i) \right| \tag{23}$$

is the test statistic. Note, that for a given value of d_n, $P(\kappa > d_n)$ is the standard Kolmogorov-Smirnov test p-value, so that the K-S test tables can be applied in the wedf approach.

4.3 Critical values

Note, that the described above testing procedure is valid only if the parameters of the hypothesized distribution are known. Unfortunately in typical applications the parameters have to be estimated beforehand. If this is the case, then the critical values for the test must be reduced (Čižek et al., 2011). In other words, if the value of the test statistics d_n is d, then the p-value is overestimated by $P(d_n \geq d)$. Hence, if this probability is small, then the p-value will be even smaller and the hypothesis will be rejected. However, if it is large then we have to obtain a more accurate estimate of the p-value.

To cope with this problem, Ross (2002) recommends to use Monte Carlo simulations. In our case the procedure reduces to the following steps. First, the parameter vector $\hat{\theta}$ is estimated from the dataset and the test statistic d_n is calculated according to formula (14). Next, $\hat{\theta}$ is used as a parameter vector for N simulated samples from the assumed model. For each sample the new parameter vector $\hat{\theta}_i$ is estimated and the new test statistic d_n^i is calculated using formula (14). Finally, the p-value is obtained as the proportion of simulated samples with the test statistic values higher or equal to d_n, i.e. p-value $= \frac{1}{N}\#\{i : d_n^i \geq d_n\}$.

5 Application to electricity spot prices

In this study we present how the techniques introduced in Section 3 can be used to efficiently calibrate MRS models to electricity spot prices and test their goodness-of-fit using the ewedf approach described in Section 4. We use mean daily (baseload)

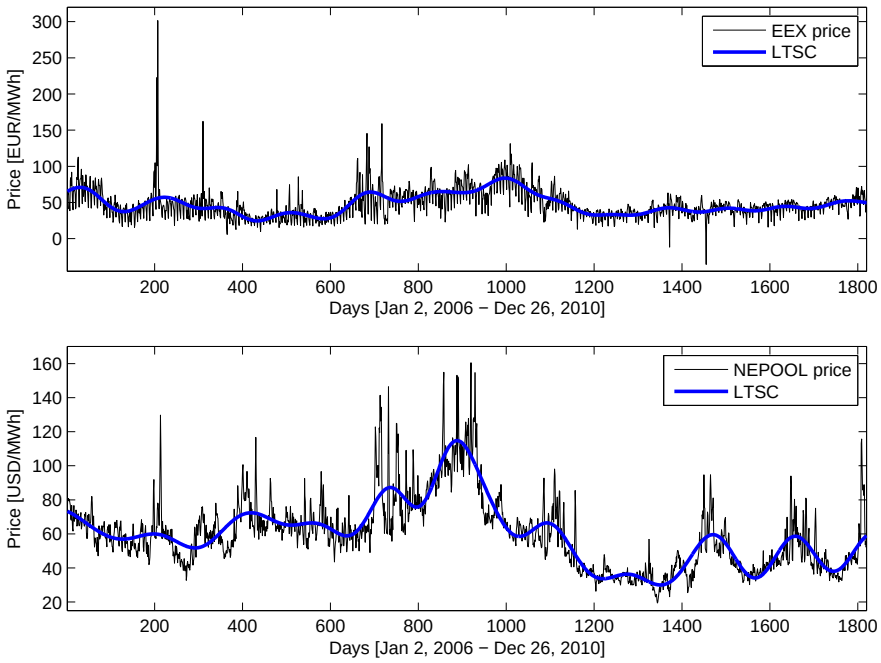

Fig. 3 Mean daily spot EEX (*top*) and NEPOOL (*bottom*) prices and the estimated long-term seasonal components (LTSC; thick blue lines).

spot prices from two major power markets: the European Energy Exchange (EEX; Germany) and the North American New England power market (NEPOOL, U.S.). 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. Both samples total 1820 daily observations (or 260 full weeks) and cover the roughly 5-year period January 2, 2006 – December 26, 2010, see Figure 3.

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, arising due to changing climate conditions, like temperature and the number of daylight hours, and business activity. These seasonal fluctuations can be then observed in electricity spot prices. In the mid- and long-term also the fuel price levels (of natural gas, oil, coal) influence electricity prices.

Not wanting to focus the paper on modeling the fundamental drivers of electricity prices, a single non-parametric long-term seasonal component (LTSC) is used here to represent the long-term non-periodic fuel price levels, the changing climate/consumption conditions throughout the years and strategic bidding practices. As shown by Janczura and Weron (2010), a wavelet-estimated LTSC pretty well reflects the 'average' fuel price level, understood as a combination of NG, crude oil and coal prices; see also Eydeland and Wolyniec (2012) and Karakatsani and Bunn

(2010) for a treatment of fundamental and behavioral drivers of electricity prices. On the other hand, as discussed recently in Janczura and Weron (2012a), the use of the wavelet-based LTSC is somewhat controversial. Predicting it beyond the next few weeks is a difficult task, because individual wavelet functions are quite localized in time or (more generally) in space. Preliminary research suggests, however, that despite this feature the wavelet-based LTSC can be extrapolated into the future yielding a better on-average prediction of the level of future spot prices than an extrapolation of a sinusoidal LTSC (Nowotarski et al., 2011). 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., 2008; Weron, 2008), which should somehow be separated from the spot price forecast for it to be useful.

In this empirical study we assume that the electricity spot price, 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 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 Figure 3.

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 mean of prices corresponding to each day of the week (the German and U.S. national holidays are treated as the eight day of the week). Finally, the deseasonalized prices, i.e. $X_t = P_t - T_t - s_t$, are shifted so that the mean of the new process X_t 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 Figures 4 and 5.

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 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.

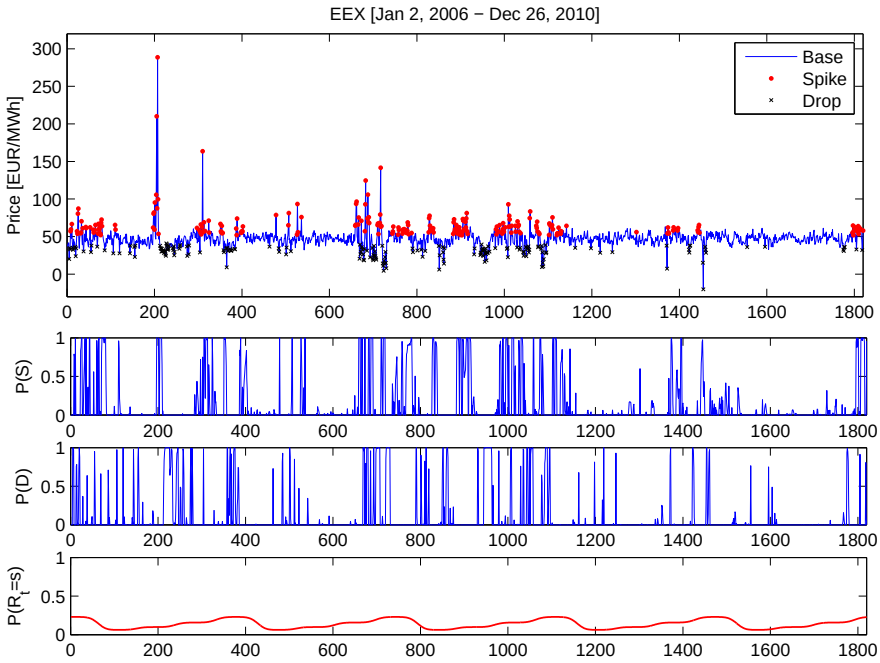

Fig. 4 Calibration results of a MRS model with three independent regimes fitted to the deseasonalized EEX prices depicted in the top panel of Fig. 3. This is the 'optimal' calibration scenario with the cutoffs q_2 and q_3 obtained as a result an optimization procedure. The lower panels display the conditional probabilities $P(S) = P(R_t = 2|x_1, x_2, ..., x_T)$ and $P(D) = P(R_t = 3|x_1, x_2, ..., x_T)$ of being in the spike or drop regime, respectively, and the time-varying unconditional probabilities $P(R_t = s)$ of being in the spike regime. The prices classified as spikes or drops, i.e. with $P(S) > 0.5$ or $P(D) > 0.5$, are denoted by dots or 'x' in the upper panel.

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:

$$X_t = \begin{cases} X_{t,1} & \text{if } R_t = 1, \\ X_{t,2} & \text{if } R_t = 2, \\ X_{t,3} & \text{if } R_t = 3. \end{cases} \tag{24}$$

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

$$X_{t,1} = \alpha_1 + (1 - \beta_1) X_{t-1,1} + \sigma_1 |X_{t-1,1}|^{\gamma_1} \varepsilon_t, \tag{25}$$

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:

$$\log(X_{t,2} - X(q_2)) \sim \mathrm{N}(\alpha_2, \sigma_2^2), \quad X_{t,2} > X(q_2). \tag{26}$$

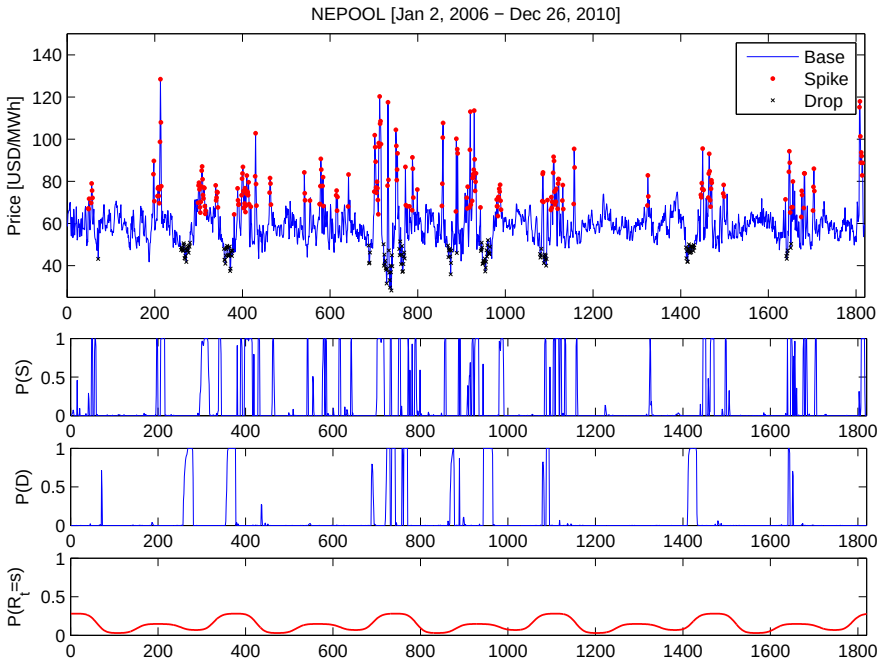

Fig. 5 Calibration results of a MRS model with three independent regimes fitted to the deseasonalized NEPOOL prices depicted in the bottom panel of Fig. 3. This is the 'optimal' calibration scenario with the cutoffs q_2 and q_3 obtained as a result an optimization procedure. The lower panels display the conditional probabilities $P(S) = P(R_t = 2|x_1, x_2, ..., x_T)$ and $P(D) = P(R_t = 3|x_1, x_2, ..., x_T)$ of being in the spike or drop regime, respectively, and the time-varying unconditional probabilities $P(R_t = s)$ of being in the spike regime. Like in Figure 4, the prices classified as spikes or drops, i.e. with $P(S) > 0.5$ or $P(D) > 0.5$, are denoted by dots or 'x' in the upper panel.

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

$$\log(-X_{t,3} + X(q_3)) \sim \mathrm{N}(\alpha_3, \sigma_3^2), \quad X_{t,3} < X(q_3). \tag{27}$$

In the above formulas $X(q_i)$ denotes the q_i-quantile, $q_i \in (0,1)$, of the dataset.

The deseasonalized prices X_t, the conditional probabilities of being in the spike $P(R_t = 2|x_1, x_2, ..., x_T)$ or drop $P(R_t = 3|x_1, x_2, ..., x_T)$ regime for the analyzed datasets and the time-varying unconditional probabilities $P(R_t = s)$ of being in the spike regime are displayed in Figures 4 and 5. The prices classified as spikes or drops, i.e. with $P(R_t = 2|x_1, x_2, ..., x_T) > 0.5$ or $P(R_t = 3|x_1, x_2, ..., x_T) > 0.5$, are additionally denoted by dots or 'x'. The estimated model parameters and the numbers of observations classified as spikes or drops are given in Table 1. The calibration results are reported for three different scenarios called: 'optimal' (with the cutoffs q_2 and q_3 obtained as a result an optimization procedure), 'quartiles' (with the cutoffs being arbitrarily set to the 1st and 3rd quartiles of the deseasonalized dataset:

Table 1 Calibration results under three different scenarios for the MRS model with three independent regimes (25)-(27) fitted to the deseasonalized EEX and NEPOOL prices. For scenario definitions see text. The numbers of observations classified as spikes (#S), i.e. with $P(R_t = 2|x_1,x_2,...,x_T) > 0.5$, or drops (#D), i.e. with $P(R_t = 3|x_1,x_2,...,x_T) > 0.5$, are additionally provided in the last two columns.

Calibration	Parameters								Probabilities				
scenario	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}	#S	#D
EEX													
optimal (0.25%, 0.69%)	18.82	0.40	0.40	0.51	2.21	0.86	2.37	0.39	0.90	0.68	0.60	238	193
quartiles (0.25%, 0.75%)	18.95	0.40	0.64	0.45	2.23	0.91	2.38	0.38	0.91	0.64	0.61	192	189
median (0.5%, 0.5%)	18.47	0.39	0.28	0.55	2.62	0.51	2.59	0.29	0.90	0.68	0.66	200	240
NEPOOL													
optimal (0.24%, 0.65%)	14.68	0.25	0.86	0.35	2.61	0.50	2.04	0.26	0.95	0.75	0.87	239	145
quartiles (0.25%, 0.75%)	15.12	0.26	3.28	0.19	2.45	0.66	2.07	0.25	0.95	0.74	0.87	229	140
median (0.5%, 0.5%)	14.84	0.25	0.33	0.47	2.84	0.33	2.55	0.10	0.95	0.76	0.87	234	161

Table 2 Goodness-of-fit statistics for the MRS model with three independent regimes (25)-(27) fitted to the deseasonalized EEX and NEPOOL prices. For parameter estimates see Table 1.

Calibration	K-S test p-values				LogL
scenario	Base	Spike	Drop	Model	
EEX					
optimal (0.25%, 0.69%)	0.8433	0.3106	0.9323	0.5887	-5432.17
quartiles (0.25%, 0.75%)	0.8912	0.1940	0.7898	0.4370	-5459.58
median (0.5%, 0.5%)	0.8857	0.0156	0.2767	0.1355	-5492.59
NEPOOL					
optimal (0.24%, 0.65%)	0.1480	0.4556	0.9529	0.3690	-5222.08
quartiles (0.25%, 0.75%)	0.1159	0.3233	0.9655	0.2339	-5233.92
median (0.5%, 0.5%)	0.1925	0.7821	0.9195	0.2988	-5232.62

$q_2 = 0.75\%$, $q_3 = 0.25\%$) and 'median' (with the cutoffs being arbitrarily set to the median of the deseasonalized dataset: $q_2 = q_3 = 0.5\%$).

Although the estimated parameters, probabilities and the numbers of identified spikes and drops differ between the scenarios, the obtained base regime parameters are consistent with the well known properties of electricity prices. In particular, $\beta_1 \in [0.25, 0.40]$ indicates a relatively high speed of mean-reversion, while positive values of $\gamma \in [0.19, 0.55]$ are responsible for the 'inverse leverage effect'. Considering probabilities p_{ii} of staying in the same regime we obtain quite high values for each of the regimes, ranging from 0.60 for the drop regime in the EEX market up to 0.95 for the base regime in the NEPOOL market. As a consequence, on average there are many consecutive observations from the same regime. Finally, since both analyzed markets are characterized by relatively similar climate conditions the patterns of spike intensity, as measured by the periodic unconditional probabilities $P(R_t = s)$, are similar (see the bottom panels in Figures 4 and 5). The spike intensity is the highest in Winter and the lowest in Spring.

In order to check the statistical adequacy of the fitted MRS models we perform 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 Section 4 and Janczura

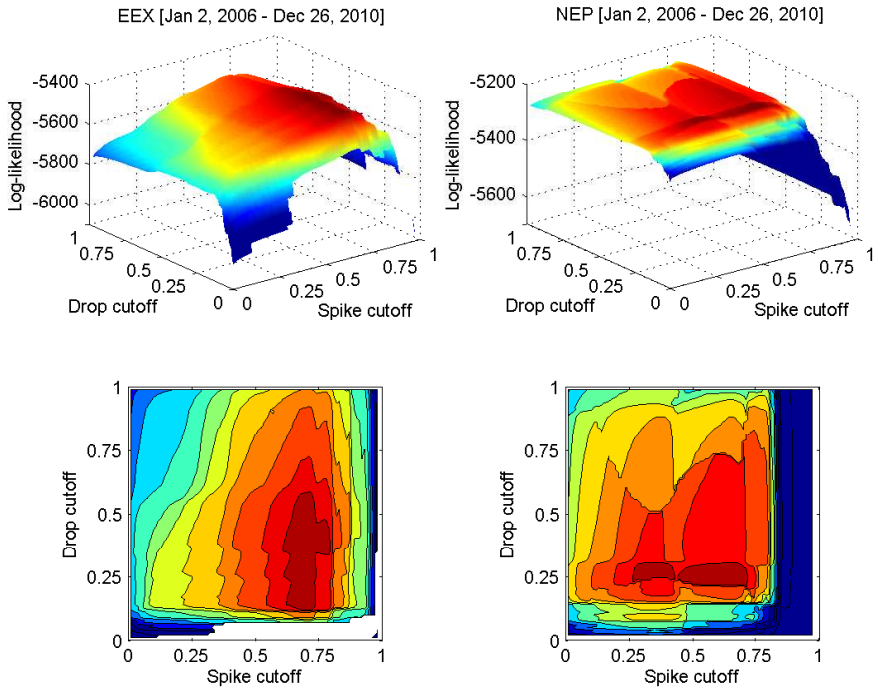

Fig. 6 Log-likelihoods of the best fitted models for each of the possible spike and drop regime cutoffs (0, 0.01, ..., 1) for EEX (*left panels*) and NEPOOL (*right panels*) deseasonalized daily prices. Bottom panels display contour plots of the upper panels. Note, that for both datasets high values for the spike cutoff and low for the drop cutoff are preferred. The optimal cutoffs for the EEX dataset are $q_2 = 0.69\%$ and $q_3 = 0.25\%$, while for the NEPOOL dataset $q_2 = 0.65\%$ and $q_3 = 0.24\%$, see Table 2.

and Weron, 2012b). The goodness-of-fit results are reported in Table 2, together with the log-likelihood values. All but one (EEX prices, spike regime, 'median' scenario) K-S test p-values are higher than the commonly used 5% significance level, hence we cannot generally reject the hypotheses that the datasets follow the fitted MRS models. Looking at the optimal cutoffs – $q_2 = 0.69\%$ and $q_3 = 0.25\%$ for the EEX dataset and $q_2 = 0.65\%$ and $q_3 = 0.24\%$ for the NEPOOL dataset – we can observe that in both cases high values for the spike cutoff and low for the drop cutoff are preferred, see also Figure 6. What is interesting, these values are relatively close to the 3rd and 1st quartiles, yet the 'quartiles' scenario is not always better than the 'median' scenario. It seems that by arbitrarily setting the cutoffs to the quartiles too many spikes in the NEPOOL dataset were excluded from being classified into the spike regime. This resulted in a relatively low log-likelihood value. Overall, we suggest to use the 'optimal' scenario as it yields significantly higher log-likelihood values and higher p-values of the K-S type test for the whole model, see Table 2.

6 Conclusions

In this paper we have reviewed the calibration and statistical validation techniques for Markov regime-switching (MRS) models of electricity spot prices. In particular, in Section 3 we have presented an efficient parameter estimation algorithm for independent regime MRS models. Instead of storing conditional probabilities for each of the possible state process paths, it 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. We have further shown how to improve the temporal fit of the models to electricity spot price data by introducing time-varying (periodic) transition probabilities and how to modify the calibration scheme by optimizing the cutoffs defining the spike and drop regimes. The latter improvement results in significantly higher log-likelihood values and higher p-values of the goodness-of-fit test for the whole model.

While most of the electricity spot price models proposed in the literature are elegant, their fit to empirical data has either been not examined thoroughly or the signs of a bad fit ignored. This can have far-reaching consequences if such misspecified models are used for forecasting or risk management applications. The goodness-of-fit tests discussed in Section 4 provide an efficient tool for accepting or rejecting a given MRS model for a particular dataset.

Finally, in Section 5 we have put the theoretical tools to use and built models of deseasonalized wholesale spot prices from the EEX and NEPOOL markets. The studied independent regime 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'. Its positive value 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, we have not checked whether the studied MRS model recovers the market observed term structure of volatility. As Janczura and Weron (2012a) have reported, the electricity forward prices implied by the considered spot price models exhibit the so-called Samuelson effect (i.e. a decrease in volatility with increasing time to maturity; for the considered models the volatility scales as $e^{-\beta(T-t)}$), but the rate of decrease is completely determined by the speed of mean-reversion β. Empirical evidence shows, however, that the rate of decrease should be large only for maturities up to a year (Kiesel et al., 2009). Perhaps, incorporating another stochastic factor would lead to a more realistic forward price curve.

Acknowledgments

This paper has benefited from conversations with the participants of the DStatG 2010 Annual Meeting, the Trondheim Summer 2011 Energy Workshop, the 2011

WPI Conference in Energy Finance, the Energy Finance Christmas Workshop (EFC11) in Wrocław and the seminars at Macquarie University, University of Sydney and University of Verona. This work was supported by funds from the National Science Centre (NCN) through grant no. 2011/01/B/HS4/01077.

References

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

Benth, F.E., Cartea, A., Kiesel, R. (2008). Pricing forward contracts in power markets by the certainty equivalence principle: Explaining the sign of the market risk premium. Journal of Banking & Finance 32(10), 2006-2021.

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

Bierbrauer, M., Menn, C., Rachev, S.T., Trück, S. (2007). Spot and derivative pricing in the EEX power market. Journal of Banking and Finance 31, 3462-3485.

Bierbrauer, M., Trück, S., Weron, R. (2004). Modeling electricity prices with regime switching models. Lecture Notes in Computer Science 3039, 859-867.

Borak, S., Weron, R. (2008). A semiparametric factor model for electricity forward curve dynamics. Journal of Energy Markets 1(3), 3-16.

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

Cartea, A., Figueroa, M., Geman, H. (2009). Modelling Electricity Prices with Forward Looking Capacity Constraints. Applied Mathematical Finance 16(2), 103-122.

Celeux, G., Durand, J.B. (2008) Selecting hidden Markov model state number with cross-validated likelihood. Computational Statistics 23, 541-564.

Cetin, M., Comert, G. (2006) Short-term traffic flow prediction with regime switching models. Transportation Research Record: Journal of the Transportation Research Board 1965, 23-31.

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

Cho, J.S., White, H. (2007) Testing for regime switching. Econometrica 75(6), 1671-1720.

Choi, S. (2009) Regime-switching univariate diffusion models of the short-term interest rate. Studies in Nonlinear Dynamics & Econometrics 13(1), Article 4.

Čižek P, Härdle W, Weron R, eds. (2011) Statistical Tools for Finance and Insurance (2nd ed.). Springer, Berlin.

D'Agostino R.B., Stevens M.A., eds. (1986) Goodness-of-fit testing techniques. Marcel Dekker, New York.

De Jong, C. (2006). The nature of power spikes: A regime-switch approach. Studies in Nonlinear Dynamics & Econometrics 10(3), Article 3.

Dempster, A., Laird, N., Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society 39, 1-38.

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

Eydeland, A., Wolyniec, K. (2012). Energy and Power Risk Management (2nd ed). Wiley, Hoboken, NJ.

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

Garcia, R. (1998) Asymptotic null distribution of the likelihood ratio test in Markov switching models. International Economic Review 39, 763-788.

Gray, S.F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. Journal of Financial Economics 42, 27-62.
Hahn, M., Frühwirth-Schnatter, S., Sass, J. (2009). Estimating models based on Markov jump processes given fragmented observation series. AStA 93, 403-425.
Hamilton, J. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57, 357-384.
Hamilton, J. (1990). Analysis of time series subject to changes in regime. Journal of Econometrics 45, 39-70.
Hamilton, J. (1996) Specification testing in Markov-switching time series models. Journal of Econometrics 70, 127-157.
Hamilton, J. (1996). Regime switching models. In: The New Palgrave Dictionary of Economics (2nd ed.).
Hirsch, G. (2009) Pricing of hourly exercisable electricity swing options using different price processes. Journal of Energy Markets 2(2), 3-46.
Hu, L., Shin, Y. (2008) Optimal test for Markov switching GARCH models. Studies in Nonlinear Dynamics & Econometrics 12(3), Article 3.
Huisman, R. (2008). The influence of temperature on spike probability in day-ahead power prices. Energy Economics 30, 2697-2704.
Huisman, R. (2009). An Introduction to Models for the Energy Markets. Risk Books.
Huisman, R., de Jong, C. (2002). Option formulas for mean-reverting power prices with spikes. ERIM Report Series Reference No. ERS-2002-96-F&A.
Huisman, R., de Jong, C. (2003). Option pricing for power prices with spikes. Energy Power Risk Management 7.11, 12-16.
Huisman, R., Mahieu, R. (2003). Regime jumps in electricity prices. Energy Economics 25, 425-434.
Janczura, J., Weron, R. (2009). Regime switching models for electricity spot prices: Introducing heteroskedastic base regime dynamics and shifted spike distributions. IEEE Conference Proceedings (EEM'09), DOI 10.1109/EEM.2009.5207175.
Janczura, J., Weron, R. (2010). An empirical comparison of alternate regime-switching models for electricity spot prices. Energy Economics 32, 1059-1073.
Janczura, J., Weron, R. (2012). Efficient estimation of Markov regime-switching models: An application to electricity spot prices. AStA - Advances in Statistical Analysis, Online First doi:10.1007/s10182-011-0181-2.
Janczura, J., Weron, R. (2012). Goodness-of-fit testing for the marginal distribution of regime-switching models. Available at MPRA: http://mpra.ub.uni-muenchen.de/36461.
Kanamura, T., Ohashi, K. (2008). On transition probabilities of regime switching in electricity prices. Energy Economics 30, 1158-1172.
Karakatsani, N.V., Bunn, D.W. (2008). Intra-day and regime-switching dynamics in electricity price formation. Energy Economics 30, 1776-1797.
Karakatsani, N.V., Bunn, D. (2010). Fundamental and behavioural drivers of electricity price volatility. Studies in Nonlinear Dynamics & Econometrics 14(4), Article 4.
Kholodnyi, V.A. (2005). Modeling power forward prices for power with spikes: A non-Markovian approach. Nonlinear Analysis 63, 958-965.
Kiesel, R., Schindlmayr, G., Börger, R.H. (2009). A two-factor model for the electricity forward market. Quantitative Finance 9(3), 279-287.
Kim, C.-J. (1994). Dynamic linear models with Markov-switching. Journal of Econometrics 60, 1-22.
Knittel, C.R., Roberts, M.R. (2005). An empirical examination of restructured electricity prices. Energy Economics 27, 791-817.
Luo, Q., Mao, X. (2007). Stochastic population dynamics under regime switching. Journal of Mathematical Analysis and Applications, 334(1), 69-84.
Mamon, R.S., Elliott, R.J., eds. (2007). Hidden Markov Models in Finance. International Series in Operations Research & Management Science, Vol. 104, Springer.

Mari, C. (2008). Random movements of power prices in competitive markets: A hybrid model approach. Journal of Energy Markets 1(2), 87-103.

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

Nowotarski, J., Tomczyk, J., Weron, R. (2011). Wavelet-based modeling and forecasting of the seasonal component of spot electricity prices. Paper presented at the Energy Finance Christmas Workshop (EFC11), Wrocław, Dec. 19-20, 2011.

Ross, S. (2002) Simulation. Academic Press, San Diego.

Schwartz, E.S., (1997). The stochastic behavior of commodity prices: Implications for valuation and hedging. Journal of Finance 52(3), 923-973.

Sen, R., Hsieh, F. (2009) A note on testing regime switching assumption based on recurrence times. Statistics and Probability Letters 79, 2443-2450.

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

Vasas, K., Eleka, P., Markusa, L. (2007) A two-state regime switching autoregressive model with an application to river flow analysis. Journal of Statistical Planning and Inference, 137(10), 3113-3126.

Weron, R. (2006). Modeling and forecasting electricity loads and prices: A statistical approach. Wiley, Chichester.

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

Weron, R. (2009). Heavy-tails and regime-switching in electricity prices. Mathematical Methods of Operations Research 69(3), 457-473.

Weron, R., Bierbrauer, M., Trück, S. (2004). Modeling electricity prices: jump diffusion and regime switching. Physica A 336, 39-48.

HSC Research Report Series 2012

For a complete list please visit http://ideas.repec.org/s/wuu/wpaper.html

01 *Inference for Markov-regime switching models of electricity spot prices* by Joanna Janczura and Rafał Weron