Goodness-of-fit testing for the marginal distribution of regime-switching models

Janczura, Joanna and Weron, Rafal

Wrocław University of Technology, Poland

10 January 2012

Working Paper, manuscript No.
(will be inserted by the editor)

Goodness-of-fit testing for the marginal distribution of regime-switching models

Joanna Janczura · Rafał Weron

This version: January 10, 2012

Abstract In this paper we propose a new goodness-of-fit testing scheme for the marginal distribution of regime-switching models. We consider models with an observable (like threshold autoregressions), as well as, a latent state process (like Markov regime-switching). The test is based on the Kolmogorov-Smirnov supremum-distance statistic and the concept of the weighted empirical distribution function. The motivation for this research comes from a recent stream of literature in energy economics concerning electricity spot price models. While the existence of distinct regimes in such data is generally unquestionable (due to the supply stack structure), the actual goodness-of-fit of the models requires statistical validation. We illustrate the proposed scheme by testing whether commonly used Markov regime-switching models fit deseasonalized electricity prices from the NEPOOL (U.S.) day-ahead market.

Keywords Regime-switching · marginal distribution · goodness-of-fit · weighted empirical distribution function · Kolmogorov-Smirnov test · conditional independence

1 Introduction

Regime-switching models have attracted a lot of attention in the recent years. A flexible specification allowing for abrupt changes in model dynamics has led to its popularity not only in econometrics (Choi, 2009; Hamilton, 2008;

Joanna 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

Rafał Weron
Institute of Organization and Management, Wrocław University of Technology, Wyb. Wyspiańskiego 27, 50-370 Wrocław, Poland
E-mail: rafal.weron@pwr.wroc.pl

Lux and Morales-Arias, 2010) but also in other as diverse fields of science as traffic modeling (Cetin and Comert, 2006), population dynamics (Luo and Mao, 2007), river flow analysis (Vasas et al., 2007) or earthquake counts (Bulla and Berzel, 2008). This paper is motivated by yet another stream of literature: electricity spot price models in energy economics (Bierbrauer et al., 2007; De Jong, 2006; Huisman and de Jong, 2003; Janczura and Weron, 2010, 2012; Karakatsani and Bunn, 2008, 2010; Mari, 2008; Misiorek et al., 2006; Weron, 2009). Regime-switching models have seen extensive use in this area due to their relative parsimony (a prerequisite in derivatives pricing) and the ability to capture the unique characteristics of electricity prices (in particular, the spiky and non-linear price behavior). 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, to our best knowledge, procedures for goodness-of-fit testing of the marginal distribution of regime-switching models have not been derived to date (with the exception of Janczura and Weron, 2009, where an ewedf-type test was introduced in the context of electricity spot price models, see Section 3.2.1 for details). With this paper we want to fill the gap. We propose an empirical distribution function (edf) based testing technique built on the Kolmogorov-Smirnov test. The procedure is readily applicable to regime-switching models with an observable, as well as, a latent state process. The derivation of the test in the latter case requires, however, a utilization of the concept of the weighted empirical distribution function (wedf).

The paper is structured as follows. In Section 2 we describe the structure of the analyzed regime-switching models and briefly explain the estimation process. In Section 3 we introduce goodness-of-fit testing procedures appropriate for regime-switching models both with observable and latent state processes. Next, in Section 4 we provide a simulation study and check the performance of the proposed technique. Since the motivation for this paper comes from the energy economics literature, in Section 5 we show how the presented testing procedure can be applied to verify the fit of Markov regime-switching models to electricity spot prices. Finally, in Section 6 we conclude.

2 Regime-switching models

2.1 Model definition

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:

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

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 HMMs, 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 $i, j = \{1, 2, ..., L\}$:

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

The current state R_t at time t depends on the past only through the most recent value R_{t-1}. 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. In general, L regime models can be considered. However, for clarity of exposition we limit the discussion in this paper to two regime models only. Note, that this is not a very restrictive limitation – at least in the context of modeling electricity spot prices – since typically two or three regimes are enough to adequately model the dynamics (Janczura and Weron, 2010; Karakatsani and Bunn, 2010). Nonetheless, all presented results are also valid for $L > 2$.

The definitions of the individual regimes can be arbitrarily chosen depending on the modeling needs. Again for the sake of clarity, in this paper we focus only on two commonly used in the energy economics literature specifications of MRS models (Ethier and Mount, 1998; De Jong, 2006; Hirsch, 2009; Huisman and de Jong, 2003; Janczura and Weron, 2010; Mari, 2008). The first one (denoted by I) assumes that the process X_t is driven by two independent regimes: (i) a mean-reverting AR(1) process:

$$X_{t,1} = \alpha + (1-\beta)X_{t-1,1} + \sigma\epsilon_t, \tag{4}$$

where the residuals ϵ_t's are independent, F^1-distributed (in the following we assume that F^1 is the standard Gaussian cdf) and (ii) an i.i.d. sample from a specified continuous, strictly monotone distribution F^2:

$$X_{t,2} \sim F^2(x), \tag{5}$$

Observe that in such a model specification the values of the first regime $X_{t,1}$ become latent when the process is in the second state. In the second specification (denoted by II) X_t is described by an AR(1) process having different parameters in each regime, namely:

$$X_t = \alpha_{R_t} + (1-\beta_{R_t})X_{t-1} + \sigma_{R_t}\epsilon_t, \quad R_t \in \{1,2\}, \tag{6}$$

where the residuals ϵ_t's are independent, $N(0,1)$-distributed random variables.

2.2 Calibration

Calibration of regime-switching models with an observable state process boils down to the problem of independently estimating parameters in each regime. In case of MRS models, though, the calibration process is not straightforward, since the state process is latent and not directly observable. We have to infer the parameters and state process values at the same time. In this paper we use a variant of the Expectation-Maximization (EM) algorithm that was first applied to MRS models by Hamilton (1990) and later refined by Kim (1994). It is a two-step iterative procedure, reaching a local maximum of the likelihood function:

- **Step 1:** Denote the observation vector by $\mathbf{x_T} = (x_1, x_2, ..., x_T)$. For a parameter vector $\theta^{(n)}$ compute the conditional probabilities $P(R_t = i|\mathbf{x_T}; \theta^{(n)})$ – the so called 'smoothed inferences' – for the process being in regime i at time t.
- **Step 2:** Calculate new and more exact maximum likelihood estimates $\theta^{(n+1)}$ using the log-likelihood function, weighted with the smoothed inferences from Step 1, i.e.

$$\log\left[L(\theta^{(n+1)})\right] = \sum_{i=1}^{2}\sum_{t=1}^{T} P(R_t = i|\mathbf{x_T}; \theta^{(n)}) \log\left[f_i(x_t|\mathbf{x_{t-1}}; \theta^{(n+1)})\right],$$

  where $f_i(x_t|\mathbf{x_{t-1}}; \theta^{(n+1)})$ is the conditional density of the i-th regime.

For a detailed description of the estimation procedure see the original paper of Kim (1994) or a recent article of Janczura and Weron (2012), where an efficient algorithm for MRS models of type I is presented.

3 Goodness-of-fit testing

In this Section we introduce a goodness-of-fit testing technique, that can be applied to evaluate the fit of regime-switching models. It is based on the Kolmogorov-Smirnov (K-S) goodness-of-fit test and verifies 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.

3.1 Testing in case of an observable state process

3.1.1 Specification I

I this case the hypothesis H_0 states that the sample $(x_1, x_2, ..., x_T)$ is generated from a regime-switching model with two independent regimes defined as: an AR(1) process (first regime) and i.i.d. F^2-distributed random variables (second regime). 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. Precisely, the following transformation is applied to each pair of consecutive AR(1) observations in regime $R_t = 1$:

$$h(x, y, k) = \frac{x - (1-\beta)^k y - \alpha \frac{1-(1-\beta)^k}{\beta}}{\sigma \sqrt{\frac{1-(1-\beta)^{2k}}{1-(1-\beta)^2}}}, \tag{7}$$

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, β and σ are the model parameters, see (4). It is straightforward to see that if H_0 is true, transformation $h(x_{t+k,1}, x_{t,1}, k)$ applied to consecutive observations from the mean-reverting AR(1) regime leads to a sample $(y_1^1, y_2^1, ..., y_{n_1}^1)$ of independent and $N(0,1)$-distributed random variables. Note, that from now on we use the following notation. The original observed sample is denoted by $(x_1, x_2, ..., x_T)$. The i.i.d. (or conditionally i.i.d. in Section 3.2) samples in each of the regimes are denoted by $(y_1^1, y_2^1, ..., y_{n_1}^1)$ and $(y_1^2, y_2^2, ..., y_{n_2}^2)$, with $n_1 + n_2 = T - 1$. Note, that for the mean-reverting regime these samples are obtained by applying transformation (7).

Further, observe that transformation $h(X_{t+k,1}, X_{t,1}, k)$ is based on subtracting the conditional mean from $X_{t+k,1}$ and standardizing it with the conditional variance. Indeed, $(1-\beta)^k X_{t,1} + \alpha \frac{1-(1-\beta)^k}{\beta}$ is the conditional expected value of $X_{t+k,1}$ given $(X_{1,1}, X_{2,1}, ..., X_{t,1})$ and $\sigma^2 \frac{1-(1-\beta)^{2k}}{1-(1-\beta)^2}$ is the respective conditional variance.

Transformation (7) 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, standard goodness-of-fit tests based on the empirical distribution function (like the Kolmogorov-Smirnov or Anderson-Darling tests, see e.g. D'Agostino and Stevens, 1986) can be applied to each of the subsamples. Recall that the Kolmogorov-Smirnov test statistic is given by:

$$D_n = \sqrt{n} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|, \tag{8}$$

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 $(y_1, y_2, ..., y_n)$, the test statistic can be calculated as

$$d_n = \sqrt{n} \max_{1 \le t \le n} \left| \sum_{k=1}^{n} \frac{1}{n} \mathbb{I}_{\{y_k \le y_t\}} - F(y_t) \right|, \tag{9}$$

where $\mathbb{I}$ is the indicator function.

The goodness-of-fit of the marginal distribution of the individual regimes can be formally tested. 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 (7), while for the second regime, 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 $Y = (F^1)^{-1}[F^2(X)]$ is F^1-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. $y_t^2 = (F^1)^{-1}[F^2(x_{t,2})]$, is i.i.d. $N(0,1)$-distributed and, hence, the testing procedure is applicable.

3.1.2 Specification II

The H_0 hypothesis now states that the sample $(x_1, x_2, ..., x_T)$ is driven by a regime-switching model defined by equation (6) with $R_t \in \{1, 2\}$. Similarly as in the independent regimes case, the testing procedure is based on extracting the residuals of the mean-reverting process. Indeed, observe that under the H_0 hypothesis the transformation $h(x_t, x_{t-1}, 1)$, defined in (7), with parameters $\alpha_{R_t}, \beta_{R_t}$ and σ_{R_t} corresponding to the current value of the state process R_t, yields an i.i.d. $N(0,1)$ distributed sample. Thus, the Kolmogorov-Smirnov test can be applied. The test statistic d_n, see (9), is calculated with the standard Gaussian cdf and the sample $(y_1, y_2, .., y_T)$ of the standardized residuals, i.e. $y_t = h(x_t, x_{t-1}, 1)$.

3.1.3 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 (9). 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 (9). 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 $= \frac{1}{N}\#\{i : d_n^i \geq d_n\}$.

3.2 Testing in case of a latent state process

3.2.1 The ewedf approach

Now, assume that the sample $(x_1, x_2, ..., x_T)$ is driven by a MRS model. The regimes are not directly observable and, hence, the standard edf approach can be used only if an identification of the state process is performed first. Recall that, as a result of the estimation procedure described in Section 2.2, 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 $P(R_t = i|x_1, x_2, ..., x_T)$. Hence, a natural choice is to relate each observation with the most probable regime by letting $R_t = i$ if $P(R_t = i|x_1, x_2, ..., x_T) > 0.5$. Then, the testing procedure described in Section 3.1 is applicable. 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'. It was introduced by Janczura and Weron (2009) in the context of electricity spot price MRS models.

3.2.2 The weighted empirical distribution function (wedf)

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 should be used in the testing procedure.

For this reason we introduce here the concept of the weighted empirical distribution function (wedf):

$$F_n(x) = \sum_{t=1}^{n} \frac{w_t \mathbb{I}_{\{y_t<x\}}}{\sum_{t=1}^{n} w_t}, \tag{10}$$

where $(y_1, y_2, ..., y_n)$ is a sample of observations and $(w_1, ..., w_n)$ are the corresponding weights, such that $0 \le w_t \le M$, $\forall_{t=1,...,n}$. It is interesting to note, that the notion of the weighted empirical distribution function appears in the literature in different contexts. Maiboroda (1996, 2000) applied it to the problem of estimation and testing for homogeneity of components of mixtures with varying coefficients. Withers and Nadarajah (2010) investigated properties of distributions of smooth functionals of $F_n(x)$. In both approaches the weights were assumed to fulfill the condition $\sum_{t=1}^{n} w_t = n$. A different choice of weights was used by Huang and Brill (2004). They proposed the level-crossing method to find weights improving efficiency of the edf in the distribution tails. Yet another approach employing the weighted distribution is the generalized (weighted) bootstrap technique, see e.g. Haeusler et al. (1991), where specified random weights are used to improve the resampling method.

However, to our best knowledge, none of the applications of wedf is related to goodness-of-fit testing of Markov regime-switching models. Here we use the wedf concept to deal with the case when observations cannot be unambiguously classified to one of the regimes and, hence, a natural choice of weights of wedf 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.

3.2.3 The wedf approach for specification II

First, let us focus on the parameter-switching specification. The H_0 hypothesis states that the sample $\mathbf{x}_T = (x_1, x_2, ..., x_T)$ is driven by the MRS model defined by equation (6). Assume that H_0 is true and the model parameters are known. Like in the observable state process case, the test cannot be applied directly to the observed sample. Let y_t^i's be the transformed variables corresponding to the i-th regime, i.e. y_t^i's are obtained as $y_t^i = [x_{t+1} - \alpha_i - (1-\beta_i)x_t]/\sigma_i$. Observe that if $R_t = i$, then y_t^i becomes the residual of the i-th regime and, hence, has the standard Gaussian distribution. The weighted empirical distribution function (wedf) is then given by:

$$F_n(x) = \frac{1}{n}\sum_{t=1}^{n}\left[P(R_t = 1|\mathbf{x}_T)\mathbb{I}_{\{y_t^1<x\}} + P(R_t = 2|\mathbf{x}_T)\mathbb{I}_{\{y_t^2<x\}}\right], \tag{11}$$

where n is the size of the sample (here $n = T - 1$). Let $\Re$ be the σ-algebra generated by the state process $\{R_t\}_{t=1,2,...,T}$, i.e. the state process history up to time T. Observe that the elements of the sum in (11) are conditionally independent given $\Re$. Indeed, if for a given t, $R_t = i$ then the t-th component of the sum becomes $\mathbb{I}_{\{y_t^i<x\}}$ and y_t^i's given $R_t = i$ form an i.i.d. $N(0,1)$-distributed sample. Moreover, the following lemma ensures that the true cdf of the residuals can be approximated by the wedf.

Lemma 1 *If H_0 is true, then F_n given by (11) is an unbiased, consistent estimator of the distribution of the residuals (in this case Gaussian).*

Note, that proofs of all lemmas and theorems formulated in this Section can be found in the Appendix.

The following theorem yields a version of the K-S test applicable to parameter-switching MRS model (6). Note, that if the state process was observable, it would boil down to the standard K-S test (Lehmann and Romano, 2005, p. 584).

Theorem 1 *Let F_n be given by (11) and F be the standard Gaussian cdf. If H_0 is true and the model parameters are known, then the statistic*

$$D_n = \sqrt{n} \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \tag{12}$$

converges (weakly) to the Kolmogorov-Smirnov distribution KS as $n \to \infty$.

If hypothesis H_0 is true then, by Theorem 1, the statistic D_n asymptotically has the Kolmogorov-Smirnov distribution. Therefore if n is large enough, the following approximation holds

$$P(D_n \geq c|H_0) \approx P(\kappa \geq c), \tag{13}$$

where $\kappa \sim KS$ and c is the critical value. Hence, the p-value for the sample $(y_1^1, y_2^1, ..., y_n^1, y_1^2, y_2^2, ..., y_n^2)$, recall that $y_t^i = [x_{t+1} - \alpha_i - (1-\beta_i)x_t]/\sigma_i$, can be approximated by $P(\kappa \geq d_n)$, where

$$d_n = \sqrt{n} \max_{1\leq t\leq n} \max_{i=1,2} \left|F_n(y_t^i) - F(y_t^i)\right| \tag{14}$$

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 easily applied in the wedf approach.

The above procedure is applicable to testing the distribution of the residuals of the (whole) model. A similar approach can be used for testing the distributions of the residuals of the individual regimes. Let the wedf for the i-th regime be defined as:

$$F_n^i(x) = \sum_{t=1}^{n} \frac{P(R_t = i|\mathbf{x}_T)\mathbb{I}_{\{y_t^i<x\}}}{\sum_{t=1}^{n} P(R_t = i|\mathbf{x}_T)}, \tag{15}$$

where again y_t^i's are the transformed variables corresponding to the i-th regime, i.e. $y_t^i = [x_{t+1} - \alpha_i - (1-\beta_i)x_t]/\sigma_i$. Further, denote the theoretical distribution of the i-th regime residuals (here Gaussian) by F^i.

Lemma 2 *If H_0 is true, then $F_n^i(x)$ given by (15) is an unbiased estimator of $F^i(x)$. Moreover, it is consistent if $\forall_{i,j=1,2}\ p_{ij} < 1$.*

An analogue of Theorem 1 can be derived.

Theorem 2 *Let F_n be given by (15) and assume that R_t is an ergodic Markov chain. If H_0 is true and the model parameters are known, then the statistic*

$$D_n^i = \sqrt{w_n} \sup_{x\in\mathbb{R}} |F_n^i(x) - F^i(x)|, \tag{16}$$

where

$$w_n = \sum_{\{i_1,i_2,...,i_n\}\in I} \mathbb{I}_{\{R_1=i_1,R_2=i_2,...,R_n=i_n\}} \frac{\left[\sum_{\{k:i_k=i\}} \mathbb{I}_{\{R_k=i\}}\right]^2}{\sum_{\{k:i_k=i\}} \mathbb{I}^2_{\{R_k=i\}}} \tag{17}$$

and $I = \{(i_1, i_2, ..., i_n) : i_k \in \{1,2\}, k = 1,2,...,n\}$ converges (weakly) to the Kolmogorov-Smirnov distribution KS as $n\to\infty$.

Observe that $\sqrt{w_n}$ can be approximated by $\frac{\sum_{t=1}^n P(R_t=i|\mathbf{x}_T)}{\sqrt{\sum_{t=1}^n P^2(R_t=i|\mathbf{x}_T)}}$. Hence, for a sample of $(y_1^i, y_2^i, ..., y_n^i)$ the test statistic is given by

$$d_n^i = \frac{\sum_{t=1}^n P(R_t=i|\mathbf{x}_T)}{\sqrt{\sum_{t=1}^n P^2(R_t=i|\mathbf{x}_T)}} \max_{1\le t\le n} \left|F_n^i(y_t^i) - F^i(y_t^i)\right| \tag{18}$$

and the standard testing procedure can be applied.

3.2.4 The wedf approach for specification I

Now, assume that the sample $(x_1, x_2, ..., x_T)$ is driven by the MRS model with independent regimes. The results of Theorems 1 and 2 can be applied, however, slight modifications of the tested sample(s) are required. First, observe that the values of the mean-reverting regime become latent, when the process is in the second state. As a consequence, the calculation of the conditional mean and variance, required for the derivation of the residuals, is not straightforward. We have:

$$\begin{aligned} E(X_{t,1}|\mathbf{x}_{t-1}) &= \alpha + (1-\beta)E(X_{t-1,1}|\mathbf{x}_{t-1}),\\ Var(X_{t,1}|\mathbf{x}_{t-1}) &= (1-\beta)^2 Var(X_{t-1,1}|\mathbf{x}_{t-1}) + \sigma^2, \end{aligned}$$

where $\mathbf{x}_{t-1} = (x_1, x_2, ..., x_{t-1})$ is the vector of preceding observations. Therefore, the standardized residuals are given by the transformation:

$$g(X_{t,1}, \mathbf{x}_{t-1}) = \frac{X_{t,1} - \alpha - (1-\beta)E(X_{t-1,1}|\mathbf{x}_{t-1})}{\sqrt{(1-\beta)^2 Var(X_{t-1,1}|\mathbf{x}_{t-1}) + \sigma^2}}, \tag{19}$$

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 equalities:

$$\begin{aligned}
E(X_{t,1}|\mathbf{x}_t) &= P(R_t = 1|\mathbf{x}_t)x_t + \\
&\quad + P(R_t \neq 1|\mathbf{x}_t)\left[\alpha + (1-\beta)E(x_{t-1,1}|\mathbf{x}_{t-1})\right], \qquad (20) \\
E(X_{t,1}^2|\mathbf{x}_t) &= P(R_t = 1|\mathbf{x}_t)x_t^2 + \\
&\quad + P(R_t \neq 1|\mathbf{x}_t)\big[\alpha^2 + 2\alpha(1-\beta)E(X_{t-1,1}|\mathbf{x}_{t-1}) + \\
&\qquad + (1-\beta)^2 E(X_{t-1,1}^2|\mathbf{x}_{t-1}) + \sigma^2\big]. \qquad (21)
\end{aligned}$$

The latter formula is a consequence of the law of iterated expectation and basic properties of conditional expected values. Finally, the values $P(R_t = 1|\mathbf{x}_t)$ are calculated from the Bayes rule during the EM estimation procedure (see e.g. Kim, 1994). Note that the transformed variables $(y_1^1, y_2^1, ..., y_{T-1}^1)$, where $y_t^1 = g(x_{t,1}, \mathbf{x}_{t-1})$, are $\Re$-independent and $N(0,1)$-distributed conditionally on $\Re$.

Now, to test the fit of the mean-reverting regime, it is enough to calculate d_n^i according to formula (18) with the standard Gaussian cdf and $y_t^1 = g(x_t, \mathbf{x}_{t-1})$. Observe, that the observations from the second regime are i.i.d. by definition, so the testing procedure is straightforward with F^2 cdf and sample $(x_1, x_2, ..., x_T)$. Moreover, the 'whole model' goodness-of-fit can be also verified. Theorem 1 is directly applicable, if the distributions of the samples corresponding to both regimes are the same $F = F^1 = F^2$. Observe that, even if $F^1 \neq F^2$, the test still can be applied using the fact that for $X \sim F^2$ we have that $Y = (F^1)^{-1}[F^2(X)]$ is F^1-distributed. The test statistic d_n is calculated as in (14) with F^1 cdf (here Gaussian) and 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, i.e. $y_t^1 = g(x_{t,1}, \mathbf{x}_{t-1})$, while $(y_1^2, y_2^2, ..., y_T^2)$ are the variables corresponding to the second regime, i.e. $y_t^2 = (F^1)^{-1}[F^2(x_t)]$.

Note, that like as in the case of an observable state process, in the wedf approach we face the problem of estimating values that are later used to compute the test statistic. Again, this problem can be circumvented with the help of Monte Carlo simulations. The p-values can be computed as the proportion of simulated MRS model trajectories with the test statistic d_n, see formulas (14) and (18), higher or equal to the value of d_n obtained from the dataset.

4 Simulations

In this Section we check the performance of the procedures introduced in Section 3.2. Due to space limitations, we focus on the more challenging case of a latent state process. To this end, we generate 10000 trajectories of each of the two MRS models, Sim #1 and Sim #2, defined in Table 1. The first model follows specification I, i.e. the first regime is driven by an AR(1) process, while the second regime is described by an i.i.d. sample of log-normally distributed

Table 1 Parameters of two MRS models analyzed in the simulation study of Section 4. The first model (Sim #1) follows specification (I), i.e. the first regime is driven by an AR(1) model, while the second regime is described by an i.i.d. sample of log-normally distributed random variables. Sim #2 is simulated from the parameter-switching AR(1) model, i.e. it follows specification (II).

	Parameters						Probabilities	
	α_1	β_1	σ_1^2	α_2	β_2	σ_2^2	p_{11}	p_{22}
Sim #1	10.0	0.8	10.0	4.0		0.5	0.9	0.2
Sim #2	1.0	0.8	1.0	3.0	0.4	0.5	0.6	0.5

random variables (Sim #1; with parameters α_2 and σ_2^2, i.e. $LN(\alpha_2, \sigma_2^2)$). Recall, that a random variable X is log-normally distributed, $LN(\alpha_2, \sigma_2^2)$, if $\log(X) \sim \mathrm{N}(\alpha_2, \sigma_2^2)$, for $X > 0$. Sim #2 is simulated from the parameter-switching AR(1) model, i.e. it follows specification II, see formula (6). The length of each trajectory is 2000 observations, which corresponds to 5.5 years of daily data. Note, that the regimes of MRS models are not directly observable and, hence, the standard edf approach cannot be used.

4.1 Known model parameters

We apply the ewedf, as well as, the wedf-based goodness-of-fit test to each simulated trajectory and then calculate the percentage of rejected hypotheses H_0 at the 5% significance level. We assume that the model parameters are known. In order to apply the test to the 'whole model' in case of Sim #1 we transform the second regime values as $[\log(X) - \alpha_2]/\sigma_2$. As a consequence we obtain a $N(0,1)$-distributed sample. The computation of $E(X_{t,1}|\mathbf{x}_t)$ in the wedf approach requires backward recursion until the previous observation from the mean-reverting regime is found, see (21). However, as the number of observations is limited, the condition $P(R_t = 1|\mathbf{x}_t) = 1$ might not be fulfilled at all. The calibration scheme requires some approximation or an additional assumption. Here we assume that for each simulated trajectory the first observation comes from the mean-reverting regime.

In the ewedf approach the tested hypothesis says that the state process is known (and coincides with the proposed classification of the observations to the regimes). As a consequence, once the regimes are identified, it is equivalent to the standard edf approach. To test how it performs for a MRS model with a latent state process we apply it to the simulated trajectories (we first identify the regimes, then test whether the sample is generated from the assumed MRS model).

The results reported in Table 2 indicate that only the wedf-based test yields correct percent of rejected hypotheses. The values obtained for the ewedf-based test are far from the expected level of 5%. The ewedf approach is more restrictive, probably due to the less accurate identification of regimes. This simple example clearly shows that in case of MRS models the wedf approach is more reliable.

Table 2 Percentage of rejected hypotheses H_0 at the 5% significance level calculated from 10000 simulated trajectories of 2000 observations of the models defined in Table 1. The results of the K-S test in the ewedf, as well as, in the wedf approach are reported independently for the two regimes (First, Second) and the whole model (Model).

	ewedf			wedf		
Regime	First	Second	Model	First	Second	Model
Sim #1	0.0569	0.8688	0.1152	0.0489	0.0470	0.0410
Sim #2	0.2196	0.0794	0.1173	0.0485	0.0501	0.0413

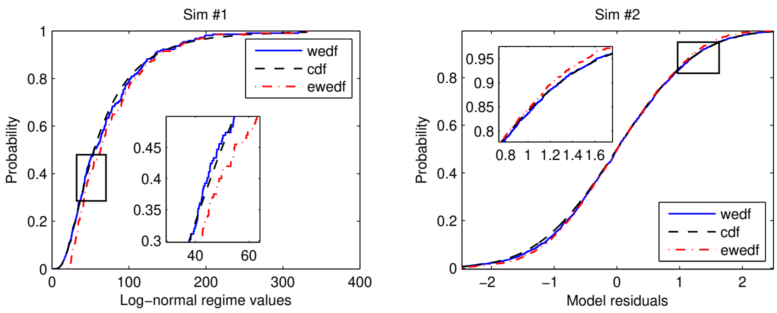

Fig. 1 Comparison of the weighted empirical distribution function (wedf), the equally-weighted empirical distribution function (ewedf) and the standard empirical distribution function (edf) calculated for a sample trajectory of a MRS model with two independent regimes – Sim #1 (left panel) and with parameter switching specification (II) – Sim #2 (right panel), see Table 1 for parameter values. The distribution functions for the model Sim #1 are plotted for the i.i.d. log-normal regime. The distribution functions for the model Sim #2 are plotted for the model residuals.

Further evidence is provided in Figure 1 where we illustrate the different types of empirical distribution functions. The wedf and ewedf functions are compared with the true edf. Note, that the edf can be calculated only when the simulated state process is known. Naturally, when dealing with real data, the state process is latent and, hence, the standard edf cannot be computed. The distribution functions are calculated separately for the log-normal regime of a sample trajectory of the MRS model Sim #1 and for the model residuals of a sample trajectory of the MRS model Sim #2, see Table 1 for parameter values. Observe that, while the wedf function replicates the true edf quite well, the ewedf approximation is not that good. This is in compliance with the rejection percentages given in Table 2.

4.2 Unknown model parameters

The simulation results presented so far were obtained with the assumption that model parameters are known. Unfortunately, in typical applications the parameters have to be estimated before the testing procedure is performed. This may result in overestimated p-values. To cope with this problem, as rec-

Table 3 Percentage of rejected hypotheses H_0 at the 5% significance level calculated from 500 simulated trajectories of 2000 observations of the models defined in Table 1 with parameters estimated from each sample. The results of the K-S test based on the ewedf, as well as, the wedf approach are reported independently for the two regimes (First, Second) and the whole model (Model). The test utilizes K-S test tables or Monte Carlo (MC) simulations with $N = 500$ repetitions. Note, that the rejection rates are only approximations based on 500 trajectories, not 10000 trajectories as in Table 2.

		ewedf			wedf		
	Regime	First	Second	Model	First	Second	Model
Sim #1	K-S test tables	0.0200	0.5900	0.0600	0.0220	0	0.0220
	MC simulation	0.0520	0.0020	0.0500	0.0520	0.0380	0.0540
Sim #2	K-S test tables	0.1340	0.0040	0.0180	0	0	0
	MC simulation	0.0280	0.0240	0.0340	0.0420	0.0340	0.0540

ommended by Ross (2002), we use Monte Carlo simulations. For a sample trajectory (of 2000 observations) simulated from model Sim #1 or Sim #2 the procedure is as follows:

- estimate the parameter vector $\hat{\theta}$ and calculate the test statistic d_n according to formula (9),
- simulate $N = 500$ trajectories with $\hat{\theta}$,
- for each trajectory estimate the new parameter vector $\hat{\theta}_i$ and calculate the new test statistic d_n^i,
- calculate p-value as the proportion of simulated trajectories with the test statistic values higher or equal to d_n, i.e. $\frac{1}{N}\#\{i : d_n^i \geq d_n\}$.

In Table 3 we report the rejection percentages at the 5% significance level. Looking at the test results based on the K-S test tables, for the ewedf approach the rejection percentages deviate significantly from the 5% level. On the other hand, for the wedf approach the p-values are overestimated, what results in rejection percentages much lower than the 5% significance level. Observe that for the model Sim #2 none of the tests were rejected. Therefore, if p-values obtained with the wedf approach are close to the significance level, the test may fail to reject a false H_0 hypothesis. This is not the case for the wedf approach with Monte Carlo simulations as the obtained rejection percentages are close to the 5% significance level. This example clearly shows that the wedf test based on the K-S test tables can only be used if it returns a p-value below the significance level (i.e. if it rejects the H_0 hypothesis) or well above the significance level. However, if the obtained p-value is close to the significance level, Monte Carlo simulations should be performed.

4.3 Power of the tests

In this Section we investigate the power of the proposed tests. To this end, for a given MRS model we simulate 500 trajectories of 100, 500 or 2000 observations each. Next, for each trajectory we calibrate a MRS model with an alternative specification of the regimes and perform goodness-of-fit tests to verify if the simulated trajectory can be driven by the alternative model. Finally, we calculate the percentages of the rejected hypotheses. The tests utilize both approaches – edf and wedf – and both methods of calculating p-values – K-S test tables and Monte Carlo simulations. We consider the following three cases:

- **AR-ARG1 vs AR-AR**: The trajectories are simulated from a MRS model defined as:

$$X_t = \alpha_{R_t} + (1-\beta_{R_t})X_{t-1} + \sigma_{R_t} X_{t-1}^{\gamma_i}\epsilon_t, \quad R_t \in \{1,2\},$$

  where $\alpha_1 = 1$, $\beta_1 = 0.8$, $\sigma_1^2 = 1$, $\gamma_1 = 0$, $\alpha_2 = 3$, $\beta_2 = 0.4$, $\sigma_2^2 = 0.05$, $\gamma_2 = 1$, $p_{11} = 0.6$ and $p_{22} = 0.5$. The model is denoted by AR-ARG1, which indicates that the first regime is driven by an AR(1) process and the second regime by a heteroskedastic autoregressive process with $\gamma = 1$ (i.e. ARG1). We test whether the simulated trajectories can be described by the model defined in equation (6), i.e. following specification II, and denoted here by AR-AR.
- **AR-E vs AR-LN**: The trajectories are simulated from a MRS model following specification I, see (4) and (5), with an exponential distribution in the second regime, i.e. $F^2 \sim \text{Exp}(\lambda)$. The model is denoted here by AR-E and its parameters are given by: $\alpha = 10$, $\beta = 0.6$, $\sigma^2 = 10$, $\lambda = 30$, $p_{11} = 0.6$ and $p_{22} = 0.5$. We test whether the simulated trajectories can be driven by a model following specification I with a log-normal distribution in the second regime (i.e. AR-LN).
- **CIR-LN vs AR-G**: The trajectories are simulated from a MRS model defined as:

$$X_{t,1} = \alpha_1 + (1-\beta_1)X_{t-1,1} + \sigma_1\sqrt{X_{t-1,1}}\epsilon_t,$$
$$X_{t,2} \sim LN(\alpha_2, \sigma_2^2),$$

  where $\alpha_1 = 1$, $\beta_1 = 0.8$, $\sigma_1^2 = 0.5$, $\alpha_2 = 2$, $\sigma_2^2 = 0.5$, $p_{11} = 0.6$ and $p_{22} = 0.5$, i.e. the first regime is a discrete time version of the square root process, also known as the CIR process (Cox et al., 1985), and the second is a log-normal random variable. Hence the name CIR-LN. We test whether the simulated trajectories can be driven by a model following specification I with a Gaussian distribution in the second regime.

The test results are summarized in Table 4. The values obtained for the individual regimes are also provided, however, as the simulated and estimated models differ, these rejection rates are highly dependent on the classification of observations to the regimes during calibration. Therefore, in the discussion that follows we focus only on the test results for the whole models. Comparing the power of the Monte Carlo approach with the one using K-S test tables, we observe that in most cases the latter method yields lower (or worse) rejection percentages. This is in compliance with the results obtained in Section 4.2. The only significant deviations from this pattern can be observed for the AR-E vs AR-LN test scenario, i.e. when the alternative model is very similar to the simulated one.

Table 4 Percentages of rejected hypotheses H_0 at the 5% significance level for the alternative models with parameters estimated for each of the 500 simulated trajectories of $T = 100$, 500 or 2000 observations. The results of the K-S test based on the ewedf, as well as, the wedf approach are reported independently for the two regimes (First, Second) and the whole model (Model). The test utilizes K-S test tables or Monte Carlo (MC) simulations with $N = 500$ repetitions. The whole model rejection rates for the best method, i.e. the wedf approach with MC simulations, are emphasised in bold.

		ewedf			wedf		
	Regime	First	Second	Model	First	Second	Model
		AR-ARG1 vs AR-AR					
T=2000	K-S test tables	0.6300	1.0000	1.0000	0.0180	1.0000	0.9960
	MC simulation	0.0540	1.0000	1.0000	0.3840	1.0000	**1.0000**
T=500	K-S test tables	0.0520	0.6520	0.4720	0.0080	0.5260	0.1080
	MC simulation	0.0860	0.9080	0.8980	0.1360	0.9820	**0.9180**
T=100	K-S test tables	0.0100	0.0200	0.0120	0	0.0040	0.0020
	MC simulation	0.0920	0.2400	0.2120	0.0860	0.3120	**0.2040**
		AR-E vs AR-LN					
T=2000	K-S test tables	0.1180	0.9980	0.9980	0.1080	0.9980	0.7700
	MC simulation	0.0560	0.9760	0.7280	0.0980	0.9980	**0.9720**
T=500	K-S test tables	0.0684	0.9738	0.5634	0.0765	0.2052	0.0604
	MC simulation	0.0825	0.1891	0.2455	0.0966	0.8873	**0.2797**
T=100	K-S test tables	0.0287	0.2480	0.0328	0.0287	0.0041	0.0184
	MC simulation	0.0389	0.0307	0.0676	0.0697	0.1701	**0.1025**
		CIR-LN vs AR-G					
T=2000	K-S test tables	1.0000	0.9491	1.0000	1.0000	0.9633	1.0000
	MC simulation	1.0000	1.0000	1.0000	1.0000	1.0000	**1.0000**
T=500	K-S test tables	1.0000	0.1566	0.9880	0.9960	0.1908	0.9940
	MC simulation	0.9980	0.7369	0.9900	1.0000	0.7771	**1.0000**
T=100	K-S test tables	0.1556	0.0182	0.0525	0.2747	0.0263	0.1455
	MC simulation	0.3071	0.1818	0.3333	0.7394	0.2182	**0.7677**

Looking at the MC simulation results obtained for the largest samples of $T = 2000$ observations, we can see that in almost all cases the false hypothesis was rejected. The lowest rejection rate for the ewedf approach was 0.7280 and for the wedf approach it was as high as 0.9720. Both were obtained for the challenging AR-E vs AR-LN test scenario. However, if the samples are smaller, the power of the tests apparently decreases. The sample size of $T = 100$ observations seems to be not enough, especially if the dynamics of the alternative models do not differ significantly, like in the AR-E vs AR-LN scenario. This is not the case if the definitions of both regimes are significantly different, as for the CIR-LN vs AR-G scenario, for which the power is satisfactory even if $T = 100$. Finally, comparing the ewedf and wedf approaches we can observe that the latter yields higher on average rejection rates.

Overall we can conclude that the power of the wedf approach with MC simulations is acceptable even for moderately sized samples covering two or more years of daily values, see the numbers in bold in Table 4. In Section 5 we will apply this test to evaluate the goodness-of-fit of two MRS models calibrated to deseasonalized electricity spot prices. The analyzed dataset comprises roughly 1800 daily observations.

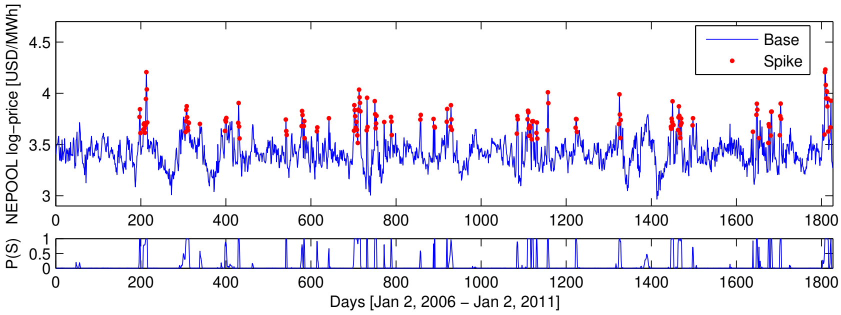

Fig. 2 Calibration results for the 2-regime MRS model with a mean reverting base regime and independent log-normally distributed ‘spikes’ fitted to NEPOOL log-prices. Observations with $P(R_t = 2|\mathbf{x}_T) > 0.5$, i.e. the ‘spikes’, are denoted by dots. The lower panel displays the probability $P(R_t = 2|\mathbf{x}_T)$ of being in the ‘spike’ regime.

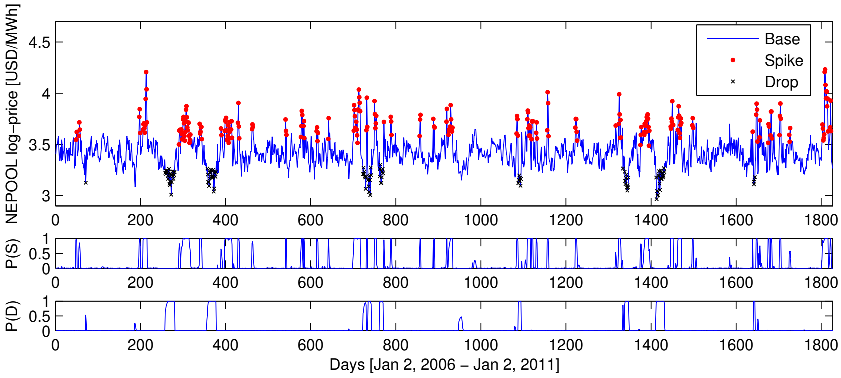

Fig. 3 Calibration results for the 3-regime MRS model with a mean reverting base regime and independent log-normally distributed ‘spikes’ and ‘drops’ fitted to NEPOOL log-prices. Observations with $P(R_t = 2|\mathbf{x}_T) > 0.5$ or $P(R_t = 3|\mathbf{x}_T) > 0.5$, i.e. the ‘spikes’ or ‘drops’, are denoted by dots or ’x’ in the upper panel. The lower panels display the probability $P(R_t = 2|\mathbf{x}_T)$ or $P(R_t = 2|\mathbf{x}_T)$ of being in the ‘spike’ or ‘drop’ regime, respectively.

5 Application to electricity spot prices

Now, we are ready to apply the new goodness-of-fit technique to electricity spot price models. We analyze the mean daily (baseload) day-ahead spot prices from the New England Power Pool SEMASS area (NEPOOL; U.S.). The sample totals 1827 daily observations (or 261 full weeks) and covers the 5-year period January 2, 2006 - January 2, 2011, see Figure 2.

It is well known that electricity spot prices exhibit several characteristic features (Eydeland and Wolyniec, 2012; Weron, 2006), which have to be taken into account when modeling such processes. These include seasonality on the annual, weekly and daily level, mean reversion and price spikes. To cope with the seasonality we use the standard time series decomposition approach and let the electricity spot price P_t be represented by a sum of two independent

Table 5 Parameters of the MRS model with mean reverting base regime and independent log-normally distributed spikes (and inverted log-normal drops in the 3-regime case) fitted to NEPOOL deseasonalized log-prices.

Model	Base regime			Spike regime		Drop regime		Probabilities		
	α	β	σ^2	α_2	s_2^2	α_3	s_3^2	p_{11}	p_{22}	p_{33}
2-regime	0.69	0.20	0.0058	-1.23	0.18	-	-	0.97	0.75	-
3-regime	0.98	0.29	0.0049	-1.38	0.21	-1.46	0.08	0.96	0.79	0.89

parts: a predictable (seasonal) component f_t and a stochastic component X_t, i.e. $P_t = f_t + X_t$. Further, to address the mean reverting and spiky behavior we let the log-prices, i.e. $Y_t = \log(X_t)$, be driven by:

– a 2-regime MRS model with mean-reverting, see (4), base regime ($R_t = 1$) and i.i.d. shifted log-normally distributed spikes ($R_t = 2$)
– or a 3-regime MRS model with mean-reverting, see (4), base regime ($R_t = 1$), i.i.d. shifted log-normally distributed spikes ($R_t = 2$) and i.i.d. drops ($R_t = 3$) distributed according to the inverted shifted log-normal law.

Recall, that X follows the shifted log-normal law (inverted shifted log-normal law) if $\log(X - q)$ (respectively $\log(q - X)$) has a Gaussian distribution. Note that q can be arbitrarily chosen, however, here for simplicity we set it to the median of the dataset.

Following Weron (2009) the deseasonalization is conducted in three steps. First, the long term seasonal component (LTSC) T_t is estimated from daily spot prices P_t using a wavelet filter-smoother of order 6 (for details see Trück et al., 2007). A single non-parametric 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), the wavelet-estimated LTSC pretty well reflects the 'average' fuel price level, understood as a combination of NG, crude oil and coal prices.

The price series without the LTSC is obtained by subtracting the T_t 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 (US 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 minimum of the new process X_t is the same as the minimum of P_t. The resulting deseasonalized time series can be seen in Figure 2. The estimated model parameters are presented in Table 5.

For both analyzed models the K-S test based on the ewedf, as well as, the wedf approach is performed. Moreover, since the standard approach based on the K-S test tables might produce overestimated p-values, the Monte Carlo results for both ewedf and wedf are also provided. Again, in order to verify the 'whole model' goodness-of-fit, we transform the spike and drop regime observations so that both samples are $N(0, 1)$-distributed. The p-values are reported in Table 6. For the 2-regime model the p-values obtained from the

Table 6 p-values of the K-S test based on the ewedf and wedf approach for both models. Values exceeding the 5% threshold are emphasized in bold.

	ewedf				wedf			
Regime	Base	Spike	Drop	Model	Base	Spike	Drop	Model
2-regime model								
K-S test tables	**0.21**	**0.27**	-	**0.29**	**0.08**	**0.93**	-	**0.13**
MC simulations	0.01	**0.11**	-	**0.07**	0.00	**0.69**	-	0.00
3-regime model								
K-S test tables	**0.56**	**0.25**	**0.98**	**0.69**	**0.38**	**0.71**	**0.92**	**0.33**
MC simulations	**0.19**	**0.06**	**0.86**	**0.42**	**0.26**	**0.25**	**0.49**	**0.15**

K-S test tables indicate that the model cannot be rejected at the 5% significance level. However, the base regime and the model p-values are still quite low, so the conclusions of the test should be verified with the Monte Carlo simulations. Indeed, for the Monte Carlo based test only the spike regime yields a satisfactory fit, as the p-value is well above the 5% significance level. The base regime, as well as, the whole model distribution can be rejected at any reasonable level. Apparently, the base regime process cannot model the sudden drops in the NEPOOL log-prices. However, if a third regime (modeling price drops) is introduced, the MRS model yields a satisfactory fit. In the 3-regime case none of the tests can be rejected at the 5% significance level.

6 Conclusions

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. As the empirical study of Section 5 has shown, even reasonably looking and popular models should be carefully tested before they are put to use in trading or risk management departments. The goodness-of-fit wedf-based test introduced in Section 3.2.2 provides an efficient tool for accepting or rejecting a given Markov regime-switching (MRS) model for a particular data set.

However, in this paper we have not restricted ourselves to MRS models but pursued a more general goal. Namely, we have proposed a goodness-of-fit testing scheme for the marginal distribution of regime-switching models, including variants with an observable and with a latent state process. For both specifications we have described the testing procedure. The models with a latent state process (i.e. MRS models) required the introduction of the concept of the weighted empirical distribution function (wedf) and a generalization of the Kolmogorov-Smirnov test to yield an efficient testing tool.

We have focused on two commonly used specifications of regime-switching models in the energy economics literature – one with dependent autoregressive states and a second one with independent autoregressive and i.i.d. regimes. Nonetheless, the proposed approach can be easily applied to other specifications of regime-switching models (for instance, to 3-regime models with heteroscedastic base regime dynamics; see Janczura and Weron, 2010). Very likely it can be also extended to other goodness-of-fit edf-type tests, like the Anderson-Darling. As the latter puts more weight to the observations in the tails of the distribution than the Kolmogorov-Smirnov test, it might be more discriminatory and provide a better testing tool for extremely spiky data. Future work will be devoted to this issue.

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.

Appendix

Proof (Lemma 1) Let

$$F_n(x) = \frac{1}{n}\sum_{t=1}^{n}\Big[P(R_t=1|\mathbf{x}_T)\mathbb{I}_{\{y_t^1<x\}} + P(R_t=2|\mathbf{x}_T)\mathbb{I}_{\{y_t^2<x\}}\Big] \tag{22}$$

and observe that

$$\begin{aligned} E[F_n(x)|\Re] &= \\ &= \frac{1}{n}\sum_{t=1}^{n}\Big(E[E(\mathbb{I}_{\{R_t=1\}}|\mathbf{x}_T)\mathbb{I}_{\{y_t^1<x\}}|\Re] + E[E(\mathbb{I}_{\{R_t=2\}}|\mathbf{x}_T)\mathbb{I}_{\{y_t^2<x\}}|\Re]\Big) = \\ &= \frac{1}{n}\sum_{t=1}^{n}\Big(\mathbb{I}_{\{R_t=1\}}E(\mathbb{I}_{\{y_t^1<x\}}|\{R_t=1\}) + \mathbb{I}_{\{R_t=2\}}E(\mathbb{I}_{\{y_t^2<x\}}|\{R_t=2\})\Big) = \\ &= \frac{1}{n}\sum_{t=1}^{n}\big(\mathbb{I}_{\{R_t=1\}}F(x) + \mathbb{I}_{\{R_t=2\}}F(x)\big) = F(x), \end{aligned} \tag{23}$$

where $F(x)$ is the distribution of the residuals and $\Re$ is the σ-algebra generated by the state process values. Similarly,

$$\begin{aligned} Var[F_n(x)|\Re] &= \\ &= \frac{1}{n^2}\sum_{t=1}^{n}\Bigg[E\Bigg\{\Big[E(\mathbb{I}_{\{R_t=1\}}|\mathbf{x}_T)\mathbb{I}_{\{y_t^1<x\}} + E(\mathbb{I}_{\{R_t=2\}}|\mathbf{x}_T)\mathbb{I}_{\{y_t^2<x\}}\Big]^2\Bigg|\Re\Bigg\} - \\ &\quad -[F(x)]^2\Bigg] = \\ &= \frac{1}{n^2}\sum_{t=1}^{n}\Big[\mathbb{I}_{\{R_t=1\}}E[\mathbb{I}^2_{\{y_t^1<x\}}|\{R_t=1\}] + \mathbb{I}_{\{R_t=2\}}E[\mathbb{I}^2_{\{y_t^2<x\}}|\{R_t=2\}] \\ &\quad -[F(x)]^2\Big] = \frac{1}{n^2}\sum_{t=1}^{n}\Big[\mathbb{I}_{\{R_t=1\}}F(x) + \mathbb{I}_{\{R_t=2\}}F(x) - [F(x)]^2\Big] = \\ &= \frac{1}{n}F(x)[1-F(x)]. \end{aligned} \tag{24}$$

Next, from the conditional Kolmogorov inequality (for details see Prakasa Rao, 2009), for any $\delta > 0$ we have (a.s.):

$$P\left(\left|F_n(x) - E[F_n(x)|\Re]\right| > \delta \middle| \Re\right) \leq \frac{Var[F_n(x)|\Re]}{\delta^2} = \tag{25}$$

$$= \frac{F(x)[1-F(x)]}{n\delta^2}. \tag{26}$$

As a consequence,

$$P(|F_n(x) - F(x)| > \delta) = E\left[P\left(\left|F_n(x) - E[F_n(x)|\Re]\right| > \delta \middle| \Re\right)\right] \leq \tag{27}$$

$$\leq \frac{F(x)[1-F(x)]}{n\delta^2} \tag{28}$$

and $F_n(x)$ converges in probability to $F(x)$ as $n \to \infty$.

Proof (Theorem 1) First, note that $F(x) \in \{0,1\}$ implies $F_n(x) = F(x)$ and $\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| = \sup_{x\in D} |F_n(x) - F(x)|$, where $D = \mathbb{R}\backslash\{x : F(x) = 0 \vee F(x) = 1\}$. Therefore in the following we will limit ourselves to the case $0 < F(x) < 1$.

Let

$$Y_t = \frac{P(R_t = 1|\mathbf{x}_T)\mathbb{I}_{\{Y_t^1<x\}} + P(R_t = 2|\mathbf{x}_T)\mathbb{I}_{\{Y_t^2<x\}}}{n},$$

where Y_t^i are the transformed variables of the i-th regime, i.e. $Y_t^i = [x_{t+1} - \alpha_i - (1-\beta_i)x_t]/\sigma_i$. Observe that, given $R_t = i$, $Y_t = \frac{1}{n}\mathbb{I}_{\{Y_t^i<x\}}$ and Y_t^i becomes the residual of the i-th regime. Therefore, $Y_1, Y_2, ..., Y_T$ are $\Re$-independent, where $\Re$ is the σ-algebra generated by the state process $\{R_t\}_{t=1,2,...,T}$, and have the same conditional distribution. Hence, from the conditional version of the Central Limit Theorem (for details see Grzenda and Zieba, 2008) we have:

$$\frac{\sum_{t=1}^n Y_t - nE[Y_t|\Re]}{\sqrt{n}\sqrt{Var(Y_t|\Re)}} \xrightarrow{d} N(0,1). \tag{29}$$

Next, note that $E(Y_t|\Re) = \frac{1}{n}F(x)$, see equation (23), and $Var(Y_t|\Re) = \frac{1}{n^2}F(x)[1-F(x)]$, see equation (24). Hence, (29) yields:

$$\frac{F_n(x) - F(x)}{\frac{1}{\sqrt{n}}\sqrt{F(x)[1-F(x)]}} \xrightarrow{d} N(0,1). \tag{30}$$

The latter is equivalent to

$$\sqrt{n}[F_n(x) - F(x)] \xrightarrow{d} W^B_{F(x)}, \tag{31}$$

where W_y^B is a Brownian bridge, i.e. $W_y^B \sim N(0, y(1-y))$, see e.g. Lehmann and Romano (2005), p. 585. Let $y = F(x)$ and observe that $\mathbb{I}_{\{y_t^i<x\}} = \mathbb{I}_{\{y_t^i<F^{-1}(y)\}} = \mathbb{I}_{\{F(y_t^i)<y\}}$. Moreover, if y_t^i are F-distributed, then $F(y_t^i)$ are driven by the uniform distribution on $[0,1]$. Formula (31) ensures that $Z_n(y) = \sqrt{n}[F_n(F^{-1}(y)) - y]$ converges pointwise to a Brownian bridge. In order to prove the convergence of $Z_n(y)$ in $D([0,1])$, i.e. in the space of right continuous functions that have left-hand limits, it is enough to show: (i) the weak convergence of the finite dimensional distributions of Z_n and that (ii)

$$E\left[|Z_n(y) - Z_n(y_1)|^\gamma |Z_n(y_2) - Z_n(y)|^\gamma\right] \leq [g(y_2) - g(y_1)]^{2\alpha} \tag{32}$$

for $y_1 < y < y_2$ and $n \geq 1$, where $\gamma \geq 0$, $\alpha > 1/2$ and g is a non-decreasing, continuous function on [0,1] (see Theorem 15.6 in Billingsley, 1968).
(i) Let $0 < y_1 < y_2 < 1$. We will show that $(Z_n(y_1), Z_n(y_2) - Z_n(y_1))$ converges weakly to $(W^B_{y_1}, W^B_{y_2} - W^B_{y_1})$. First, observe that $(Z_n(y_1), Z_n(y_2) - Z_n(y_1))$ conditional on $\Re$ is multinomially distributed with variances $(y_1(1-y_1), (y_2-y_1)[1-(y_2-y_1)])$ and covariance

$-y_1(y_2 - y_1)$. This can be calculated using the same arguments as in the proof of Lemma 1. Hence, by the central limit theorem for multinomial trials, as $n \to \infty$, for any $s \in \mathbb{R}^2$ we have

$$\varphi^{\Re}_{(Z_n(y_1), Z_n(y_2)-Z_n(y_1))}(s) \to \varphi_{(W^B_{y_1}, W^B_{y_2}-W^B_{y_1})}(s) \tag{33}$$

where φ_Z is the characteristic function of Z and $\varphi^{\Re}_Z$ is the conditional (on $\Re$) characteristic function of Z. Finally, by the dominated convergence theorem, $(Z_n(y_1), Z_n(y_2) - Z_n(y_2))$ converges weakly to $(W_{y_1}, W_{y_2} - W_{y_1})$. The convergence of finite dimensional distributions for any $0 < y_1 < y_2 < ... < y_m < 1$ follows with the same arguments.
(ii) In order to prove that condition (32) is fulfilled, observe that, given $\{R_t\}_{t=1,2,...,T}$, the wedf F_n becomes the standard edf and, hence, $Z_n(y)$ is a standard empirical process. Thus, we have (for a proof see Billingsley, 1968)

$$E\left[|Z_n(y) - Z_n(y_1)|^\gamma |Z_n(y_2) - Z_n(y)|^\gamma |\, \Re\right] \le [g(y_2) - g(y_1)]^{2\alpha}, \tag{34}$$

what obviously implies (32).
Finally, by (i) and (ii) $Z_n(y)$ converges to $W^B(y)$ in $D([0,1])$ and by the continuous mapping theorem we obtain that

$$\sqrt{n} \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| = \sup_{y\in[0,1]} |Z_n(y)| \xrightarrow{d} KS, \tag{35}$$

where KS is the Kolmogorov-Smirnov distribution, i.e. a distribution of $\sup_{0\le y\le 1} |W^B_y|$.

Proof (Lemma 2) Let $I = \{(i_1, i_2, ..., i_n) : i_k \in \{1,2\}\}$ and

$$F^i_n(x) = \sum_{t=1}^{n} \frac{P(R_t = i|\mathbf{x}_T)\mathbb{I}_{\{y^i_t < x\}}}{\sum_{t=1}^n P(R_t = i|\mathbf{x}_T)}. \tag{36}$$

We have

$$\begin{aligned}
&E[F^i_n(x)|\Re] = &(37)\\
&= \sum_{t=1}^{n} \left(E\left[\frac{E(\mathbb{I}_{\{R_t=i\}}|\mathbf{x}_T)\mathbb{I}_{\{y^i_t<x\}}}{\sum_{t=1}^n E(\mathbb{I}_{\{R_t=i\}}|\mathbf{x}_T)} \middle| \Re \right] \right) = \\
&= \sum_{t=1}^{n} \sum_{\{i_1,i_2,...,i_n\}\in I} \mathbb{I}_{\{R_1=i_1,R_2=i_2,...,R_n=i_n\}} \frac{\mathbb{I}_{\{R_t=i\}}}{\sum_{\{k:i_k=i\}} \mathbb{I}_{\{R_k=i\}}} E(\mathbb{I}_{\{y^i_t<x\}}|\{R_t = i\}) = \\
&= F(x) \sum_{\{i_1,i_2,...,i_n\}\in I} \mathbb{I}_{\{R_1=i_1,R_2=i_2,...,R_n=i_n\}} \frac{\sum_{\{k:i_k=i\}} \mathbb{I}_{\{R_k=i\}}}{\sum_{\{k:i_k=i\}} \mathbb{I}_{\{R_k=i\}}} = \\
&= F^i(x) \sum_{\{i_1,i_2,...,i_n\}\in I} \mathbb{I}_{\{R_1=i_1,R_2=i_2,...,R_n=i_n\}} = F^i(x),
\end{aligned}$$

where $\Re$ is the σ-algebra generated by the state process values.
Similarly,

$$\begin{aligned}
&Var[F^i_n(x)|\Re] = &(38)\\
&= \sum_{t=1}^{n} \left\{ \sum_{\{i_1,i_2,...,i_n\}\in I} \mathbb{I}_{\{R_1=i_1,R_2=i_2,...,R_n=i_n\}} \frac{\mathbb{I}^2_{\{R_t=i\}}}{\left[\sum_{\{k:i_k=i\}} \mathbb{I}_{\{R_k=i\}}\right]^2} F^i(x) - \right.\\
&\left. -[F^i(x)]^2 \right\} =
\end{aligned}$$

$$= \sum_{t=1}^{n} \left\{ \sum_{\{i_1,i_2,\ldots,i_n\}\in I} \mathbb{I}_{\{R_1=i_1,R_2=i_2,\ldots,R_n=i_n\}} \frac{\mathbb{I}^2_{\{R_t=i\}}}{\left[\sum_{\{k:i_k=i\}} \mathbb{I}_{\{R_k=i\}}\right]^2} F^i(x) - \right.$$

$$\left. - \sum_{\{i_1,i_2,\ldots,i_n\}\in I} \mathbb{I}_{\{R_1=i_1,R_2=i_2,\ldots,R_n=i_n\}} \frac{\mathbb{I}^2_{\{R_t=i\}}}{\left[\sum_{\{k:i_k=i\}} \mathbb{I}_{\{R_k=i\}}\right]^2} [F^i(x)]^2 \right\} =$$

$$= F^i(x)[1-F^i(x)] \sum_{\{i_1,i_2,\ldots,i_n\}\in I} \mathbb{I}_{\{R_1=i_1,R_2=i_2,\ldots,R_n=i_n\}} \frac{\sum_{\{k:i_k=i\}} \mathbb{I}^2_{\{R_k=i\}}}{\left[\sum_{\{k:i_k=i\}} \mathbb{I}_{\{R_k=i\}}\right]^2}.$$

Using the conditional Kolmogorov inequality we obtain (a.s.):

$$P\left(\left|F_n^i(x) - E[F_n^i(x)|\Re]\right| > \delta \middle| \Re\right) \leq \tag{39}$$

$$\leq \frac{F^i(x)[1-F^i(x)]}{\delta^2} \sum_{\{i_1,i_2,\ldots,i_n\}\in I} \mathbb{I}_{\{R_1=i_1,R_2=i_2,\ldots,R_n=i_n\}} \frac{\sum_{\{k:i_k=i\}} \mathbb{I}^2_{\{R_k=i\}}}{\left[\sum_{\{k:i_k=i\}} \mathbb{I}_{\{R_k=i\}}\right]^2}.$$

Now, taking the expected value and using the fact that $E[F_n^i(x)|\Re] = F^i(x)$ we have

$$P(|F_n^i(x) - F^i(x)| > \delta) \leq \tag{40}$$

$$\leq \frac{F^i(x)[1-F^i(x)]}{\delta^2} E\left(\sum_{\{i_1,i_2,\ldots,i_n\}\in I} \mathbb{I}_{\{R_1=i_1,R_2=i_2,\ldots,R_n=i_n\}} \frac{\sum_{\{k:i_k=i\}} \mathbb{I}^2_{\{R_k=i\}}}{\left[\sum_{\{k:i_k=i\}} \mathbb{I}_{\{R_k=i\}}\right]^2} \right) \leq$$

$$\leq \frac{F^i(x)[1-F^i(x)]}{\delta^2} \left[\max_{i,j=1,2}(p_{ij})\right]^n \left[\binom{n}{0}\frac{1}{n} + \binom{n-1}{1}\frac{1}{n-1} + \ldots + \binom{n}{n-1} 1\right] \leq$$

$$\leq \frac{F^i(x)[1-F^i(x)]}{\delta^2} q^n n^2.$$

for $q = [\max_{i,j=1,2}(p_{ij})]^n$. If $q < 1$, then $F_n(x)$ converges in probability to $F(x)$ for $n \to \infty$.

Proof (Theorem 2) Let

$$Y_t^j = \frac{P(R_t = j|\mathbf{x}_T)\mathbb{I}_{\{y_t^j < x\}}}{\sum_{t=1}^n P(R_t = j|\mathbf{x}_T)}.$$

Now, observe that the conditional characteristic function of $\zeta = \frac{F_n^j(x) - E[F_n^j(x)|\Re]}{\sqrt{Var[F_n^j(x)|\Re]}}$ is given by:

$$\varphi_\zeta^\Re(s) = E\left[\exp\left(is\frac{\sum_{t=1}^n Y_t - E(\sum_{t=1}^n Y_t|\Re)}{\sqrt{Var(F_n^j(x)|\Re)}}\right)\middle|\Re\right] =$$

$$= E\left[\prod_{t=1}^n \exp\left(is\frac{Y_t - E(Y_t|\Re)}{\sqrt{Var(F_n(x)|\Re)}}\right)\middle|\Re\right], \tag{41}$$

where $\Re$ is the σ-algebra generated by the state process values.

To focus attention, assume that there were m observations from regime j, i.e. $\sum_{t=1}^n \mathbb{I}_{\{R_t=j\}} = m$ and for notational convenience set $R_t = j\, \forall_{t\in\{t_1,t_2,\ldots,t_m\}}$. Consequently, $Var(F_n^j(x)|\Re) = mVar(Y_{t_l}|\Re)$, $l = 1, 2, ..., m$. Hence, taking a time series expansion of the exponent we obtain

$$\varphi_\zeta^\Re(s) = \tag{42}$$

$$= \prod_{t=t_1}^{t_m} E\left(1 + is\frac{Y_t - E(Y_t|\Re)}{\sqrt{mVar(Y_t|\Re)}} - s^2\frac{[Y_t - E(Y_t|\Re)]^2}{2mVar(Y_t|\Re)} + o\left[\frac{s^2}{mVar(Y_t|\Re)}\right]\middle|\Re\right) =$$

$$= \left\{1 - \frac{s^2}{2m} + E\left(o\left[\frac{s^2}{mVar(Y_t|\Re)}\right]\middle|\Re\right)\right\}^m.$$

Since for an ergodic Markov chain $\sum_{t=1}^{n} \mathbb{I}_{\{R_t=j\}} \to \infty$ a.s. as $n \to \infty$ and for a fixed s, $\lim_{m\to\infty} E\left(o\left[\frac{s^2}{mVar(Y_t|\Re)}\right]\Big|\Re\right) = 0$ a.s., we obtain that

$$\varphi_{\zeta}^{\Re}(s) \to e^{-\frac{s^2}{2}} \quad \text{a.s. with } n \to \infty. \tag{43}$$

Moreover, from (37) and (38) we have

$$\frac{F_n^j(x) - E[F_n^j(x)|\Re]}{\sqrt{Var[F_n^j(x)|\Re]}} = w_n \frac{F_n^j(x) - F^j(x)}{F^j(x)[1 - F^j(x)]}, \tag{44}$$

where

$$w_n = \sum_{\{i_1, i_2, \ldots, i_n\} \in I} \mathbb{I}_{\{R_1=i_1, R_2=i_2, \ldots, R_n=i_n\}} \frac{\left[\sum_{\{k: i_k=i\}} \mathbb{I}_{\{R_k=i\}}\right]^2}{\sum_{\{k: i_k=i\}} \mathbb{I}^2_{\{R_k=i\}}}.$$

Finally, from (43) and (44) we have

$$w_n[F_n(x) - F(x)] \xrightarrow{d} N(0, F(x)[1 - F(x)]) \tag{45}$$

and the proof is completed with the same arguments as in the proof of Theorem 1.

References

Bally, V., Talay, D., (1996) The law of the Euler scheme for stochastic differential equations: II. Convergence rate of the density. Monte Carlo Methods and Applications 2, 93-128.

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.

Billingsley, P. (1968) Convergence of Probability Measures, Wiley, New York.

Billingsley, P. (1986) Probability and Measure (2nd ed.), Wiley, New York.

Brockwell, P.J., Davis, R.A. (1996) Introduction to Time Series and Forecasting (2nd ed.), Springer-Verlag, New York.

Bulla, J., Berzel, A. (2008) Computational issues in parameter estimation for stationary hidden Markov models. Computational Statistics 23: 1-18.

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

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.

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.

Cox, J.C., Ingersoll, J.E., Ross, S.A. (1985). A theory of the term structure of interest rates. Econometrica 53, 385-407.

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.

Ethier, R., Mount, T., (1998). Estimating the volatility of spot prices in restructured electricity markets and the implications for option values. PSerc Working Paper 98-31.

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

Garcia, R. (1998) Asymptotic null distribution of the likelihood ratio test in Markov switching models. International Economic Review 39, 763-788.
Grzenda, W., Zieba, W. (2008) Conditional Central Limit Theorem. International Mathematical Forum 3(31), 1521-1528.
Haeusler, E., Mason, D.M., Newton, M.A. (1991) Weighted bootstrapping of the means. CWI Quaterly, 213-228.
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.
Huang, M.L., Brill, P.H. (2004) A distribution estimation method based on level crossings. Journal of Statistical Planning and Inference 124, 45-62.
Huisman, R., de Jong, C. (2003). Option pricing for power prices with spikes. Energy Power Risk Management 7.11, 12-16.
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. Available at MPRA: http://mpra.ub.uni-muenchen.de/18784.
Janczura, J., Weron, R. (2010) An empirical comparison of alternate regime-switching models for electricity spot prices. Energy Economics 32(5), 1059-1073.
Janczura, J., Weron, R. (2012) Efficient estimation of Markov regime-switching models: An application to electricity wholesale market prices. AStA – Advances in Statistical Analysis, Online First doi:10.1007/s10182-011-0181-2.
Karakatsani, N.V., Bunn, D. (2008) Forecasting electricity prices: The impact of fundamentals and time-varying coefficients. International Journal of Forecasting 24, 764-785.
Karakatsani, N.V., Bunn, D. (2010) Fundamental and behavioural drivers of electricity price volatility. Studies in Nonlinear Dynamics & Econometrics 14(4), Article 4.
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.
Lehmann, E.L., Romano, J.P. (2005) Testing Statistical Hypotheses (3rd ed.) Springer, New York.
Luo, Q., Mao, X. (2007) Stochastic population dynamics under regime switching. Journal of Mathematical Analysis and Applications, 334(1), 69-84.
Lux, T., Morales-Arias, L. (2010) Forecasting volatility under fractality, regime-switching, long memory and student-t innovations. Computational Statistics and Data Analysis 54, 2676-2692.
Maiboroda, R.E. (1996) Estimates for distributions of components of mixtures with varying concentrations. Ukrainian Mathematical Journal 48(4), 618-622.
Maiboroda, R.E. (2000) A test for the homogeneity of mixtures with varying concentrations. Ukrainian Mathematical Journal 52(8), 1256-1263.
Mari, C. (2008). Random movements of power prices in competitive markets: A hybrid model approach. Journal of Energy Markets 1(2), 87-103.
Misiorek, A., Trueck S., Weron R. (2006) Point and interval forecasting of spot electricity prices: linear vs. non-linear time series models. Studies in Nonlinear Dynamics & Econometrics 10(3), Article 2.
Prakasa Rao, B.L.S. (2009) Conditional independence, conditional mixing and conditional association. Annals of the Institute of Statistical Mathematics 61, 441-460.
Ross, S. (2002) Simulation. Academic Press, San Diego.
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. (2009) Heavy-tails and regime-switching in electricity prices. Mathematical Methods of Operations Research 69(3), 457-473.

Withers, C.S., Nadarajah, S. (2010) The distribution and quantiles of functionals of weighted empirical distributions when observations have different distributions. arXiv:1002.4338v1.