



Munich Personal RePEc Archive

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

Janczura, Joanna and Weron, Rafal

Wrocław University of Technology, Poland

9 July 2011

Online at <https://mpra.ub.uni-muenchen.de/32532/>  
MPRA Paper No. 32532, posted 01 Aug 2011 19:31 UTC

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

Joanna Janczura · Rafał Weron

Received: July 9, 2011

**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 a commonly used Markov regime-switching model fits 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

## 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; 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,

---

Joanna Janczura  
Hugo Steinhaus Center for Stochastic Methods, 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

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; Karakatsani and Bunn, 2008, 2010; Mari, 2008; Misiolek 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, 2003; 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, see Section 3.2.1, was introduced in the context of electricity spot price models). With this paper we want to fill the gap. We propose an empirical distribution function (edf) based testing technique build on the Kolmogorov-Smirnov test. The procedure is developed for regime-switching models with an observable, as well as, a latent state process. The latter involves 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  $t$ , driven by a state process  $R_t$ :

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

The 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  $i, j = \{1, 2, \dots, L\}$ :

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

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

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

where  $\mathbf{P}'$  denotes the transpose of  $\mathbf{P}$  and  $e_i$  denotes the  $i$ th 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 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

(I) assumes that the process  $X_t$  is driven by two independent regimes: (i) a mean-reverting AR(1) process:

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

where  $\epsilon_t \sim N(0, 1)$  and (ii) an i.i.d. sample from a specified distribution  $F_2$ :

$$X_{t,2} \sim F_2(x). \quad (5)$$

In the second specification (II)  $X_t$  is described by an AR(1) process having different parameters in each regime, namely:

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

## 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 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** For a parameter vector  $\theta$  compute the conditional probabilities  $P(R_t = j | x_1, \dots, x_T; \theta)$  – the so called ‘smoothed inferences’ – for the process being in regime  $j$  at time  $t$ .
- **Step 2** Calculate new and more exact maximum likelihood estimates of  $\theta$  using the likelihood function, weighted with the smoothed inferences from step 1.

Note, that the introduction of independent regimes – like in specification (I) – results in a significantly increased computational burden. See Janczura and Weron (2011) for an efficient modification of the algorithm to overcome this problem.

## 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 i.i.d. observations come from the distribution specified by the model cannot be rejected. The procedure can be easily adapted to other empirical distribution function (edf) type tests, e.g. Anderson-Darling.

### 3.1 Testing in case of an observable state process

First, we focus on the independent regimes specification (I). The hypothesis  $H_0$  states that the sample  $(x_1, x_2, \dots, x_T)$  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  $R_t = 1$ :

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

where  $(k - 1)$  is the number of latent observations from the mean reverting regime and  $\alpha$ ,  $\beta$  and  $\sigma$  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 of independent and  $N(0, 1)$  distributed random variables.

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  $(X_{1,1}, X_{2,1}, \dots, X_{t,1})$  and  $\sigma^2 \frac{1 - (1 - \beta)^{2k}}{1 - (1 - \beta)^2}$  is the respective conditional variance.

A similar transformation can be applied to a more general class of processes given by the stochastic differential equation (SDE) of the form

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t, \quad (8)$$

where  $\mu(X_t)$  and  $\sigma(X_t)$  are deterministic functions of the current (price) level  $X_t$  and  $W_t$  is a Wiener process. Processes of this form have been recently used to build regime-switching models for interest rates (Choi, 2009) and electricity prices (Janczura and Weron, 2009). Also note, that the AR(1) process defined by (4) is a discrete time version of the SDE (8) with  $\mu(X_t) = \alpha - \beta X_t$  and  $\sigma(X_t) = \sigma$ , which is known as the Vasicek model in the fixed income literature.

Using the Euler scheme and rearranging terms of formula (8), we get that

$$\epsilon_{t-\Delta t, t} = \frac{X_t - X_{t-\Delta t} - \mu(X_{t-\Delta t})\Delta t}{\sqrt{\Delta t}\sigma(X_{t-\Delta t})} \quad (9)$$

has the standard Gaussian distribution. However, since the Euler scheme is an approximation of a continuous process, (9) is valid only for small  $\Delta t$  (for details

on errors of the Euler scheme see e.g. Bally and Talay, 1996). In contrast, transformation (7) is exact.

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:

$$D_n = \sqrt{n} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|, \quad (10)$$

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

$$d_n = \sqrt{n} \max_{1 \leq t \leq n} \left| \sum_{k=1}^n \frac{1}{n} \mathbb{I}_{\{y_k \leq y_t\}} - F(y_t) \right|, \quad (11)$$

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

The goodness-of-fit of the marginal distribution of the individual regimes, as well as, of the whole model can be formally tested. For the mean-reverting regime  $F$  is the standard Gaussian cdf and  $(y_1, y_2, \dots, y_n)$  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  $(y_1, y_2, \dots, y_n)$  is the subsample of respective observations. The ‘whole model’ testing technique is now straightforward. Indeed, observe that combining both subsamples yields an i.i.d. sample coming from a distribution being a mixture of normal and model-specified ( $F_2$ ) laws. The cdf is then given by

$$F(x) = \sum_{i=1}^2 P(R = i) F_i(x), \quad (12)$$

where  $P(R = i)$  is the probability of the process being in regime  $i$  and  $F_i$  is the cdf related to regime  $i$ . Recall, that for the mean-reverting AR(1) regime  $F_1$  is the standard Gaussian cdf. Hence, statistic (11) can be calculated with  $F$  given by (12) and the sample  $(y_1^1, y_2^1, \dots, y_{n_1}^1, y_1^2, y_2^2, \dots, y_{n_2}^2)$ , where  $(y_1^1, y_2^1, \dots, y_{n_1}^1)$  is the subsample of the standardized residuals given by (7),  $(y_1^2, y_2^2, \dots, y_{n_2}^2)$  are the observed values and  $n_1 + n_2 = T - 1$ .

Now, we focus on the case when the model dynamics is described by an AR(1) process with different parameters in each regime (specification II). The  $H_0$  hypothesis states that the sample  $(x_1, x_2, \dots, x_T)$  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  $\sigma_{R_t}$  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 now be applied. The test statistic  $d_n$ , see (11), is calculated with the standard Gaussian cdf and the sample  $(y_1, y_2, \dots, y_T)$  of the standardized residuals, i.e.  $y_t = h(x_t, x_{t-1}, 1)$ .

Note, however, that the described 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 (Čížek 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 boils down 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 (11). Next,  $\hat{\theta}$  is used as a parameter vector for  $N$  simulated samples from the assumed process. For each sample the new parameter vector  $\hat{\theta}_i$  is estimated and the new test statistic  $d_n^i$  is calculated (like in formula (11)). 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.  $\frac{1}{N} \#\{i : d_n^i \geq d_n\}$ .

## 3.2 Testing in case of a latent state process

### 3.2.1 The ewedf approach

Now, assume that the sample  $(x_1, x_2, \dots, x_T)$  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  $\tilde{P}(R_t = i) = P(R_t = i | x_1, x_2, \dots, x_T)$ . Hence, a natural choice is to relate each observation with the most probable regime by letting  $R_t = i$  if  $\tilde{P}(R_t = i) > 0.5$ . Then, the testing procedure described in the previous Section is applicable. However, we have to mention, that the hypothesis  $H_0$  now states that  $(x_1, x_2, \dots, x_T)$  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 wedf approach

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):

$$F_n(x) = \sum_{t=1}^n \frac{w_t \mathbb{I}_{\{y_t < x\}}}{\sum_{t=1}^n w_t}, \quad (13)$$

where  $y_1, y_2, \dots, y_n$  is a sample of observations and  $w_1, \dots, w_n$  are the corresponding weights, such that  $0 \leq w_t \leq M$ ,  $\forall_{t=1, \dots, n}$ . 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 weights were assumed to fulfill the condition  $\sum_{t=1}^n w_t = n$ . Different choice of weights was given 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 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. The only restrictions imposed on the choice of the weights are the ones guarantying that  $F_n(x)$  is an unbiased and consistent estimator of  $F(x)$ , namely  $\forall_{t \in \mathbb{N}} 0 \leq w_t \leq M$  and  $\lim_{n \rightarrow \infty} \sum_{t=1}^n w_t = \infty$ . To see this, note that from the Chebyshev's inequality (Billingsley, 1986, p. 65), for any  $\delta > 0$  we have

$$\begin{aligned} P(|F_n(x) - E[F_n(x)]| > \delta) &\leq \frac{\text{Var}[F_n(x)]}{\delta^2} = \frac{F(x)[1 - F(x)] \sum_{t=1}^n w_t^2}{\delta^2 (\sum_{t=1}^n w_t)^2} \leq \\ &\leq \frac{F(x)[1 - F(x)] \sum_{t=1}^n M w_t}{\delta^2 (\sum_{t=1}^n w_t)^2} = \\ &= \frac{F(x)[1 - F(x)]M}{\delta^2 \sum_{t=1}^n w_t}. \end{aligned}$$

Hence, since  $E[F_n(x)] = F(x)$ ,  $F_n(x)$  converges in probability to  $F(x)$ .

The following lemma yields a generalization of the K-S test to the case of the wedf (for a proof see the Appendix). Note, that if each  $w_t \equiv 1$ , it simplifies to the result for the standard Kolmogorov-Smirnov test (Lehmann and Romano, 2005, p. 584).

**Lemma 1** *If  $Y_1, Y_2, \dots$  are independent, generated from a continuous and strictly monotone distribution  $F$ ,  $\forall t \in \mathbb{N} \text{ Var}(Y_t) < \infty, 0 \leq w_t \leq M$  and  $\lim_{n \rightarrow \infty} \sum_{t=1}^n w_t^2 = \infty$  then the statistic*

$$D_n = \frac{\sum_{t=1}^n w_t}{\sqrt{\sum_{t=1}^n w_t^2}} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \quad (14)$$

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

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

$$P(D_n \geq c | H_0) \approx P(\kappa \geq c), \quad (15)$$

where  $\kappa \sim KS$  and  $c$  is the critical value. Hence, the  $p$ -value for the analyzed sample  $(y_1, y_2, \dots, y_n)$  can be approximated by  $P(\kappa \geq d_n)$ , where

$$d_n = \frac{\sum_{t=1}^n w_t}{\sqrt{\sum_{t=1}^n w_t^2}} \max_{1 \leq t \leq n} |F_n(y_t) - F(y_t)| \quad (16)$$

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 the K-S test tables can be easily applied in the wedf approach.

Lemma 1 can be used in case of MRS models. Indeed, if the state process  $R_t$  is a Markov chain with no transient states and  $w_t = P(R_t = j)$ , the assumptions of Lemma 1 are satisfied. Goodness-of-fit of the individual regimes, as well as, of the whole model can be verified. Again, to ensure the i.i.d. condition for the tested sample, the mean-reverting regime observations are substituted with the corresponding residuals.

First, let us focus on the parameter-switching specification (II). The  $H_0$  hypothesis states that the sample  $(x_1, x_2, \dots, x_T)$  is driven by the MRS model defined by equation (6). Observe, that the residuals of the  $i$ th regime are then obtained from the transformation  $h(x_t, x_{t-1}, 1)$ , see (7), with parameters  $\alpha_i, \beta_i$  and  $\sigma_i$ . Hence, the  $i$ th regime test statistic  $d_n$ , see (16), is calculated with the standard Gaussian cdf, the sample  $(y_1, y_2, \dots, y_{T-1})$ , where  $y_t = h(x_{t+1}, x_t, 1)$ , and the weights  $w_t = \tilde{P}(R_t = i)$ . The goodness-of-fit of the whole model can be verified based on the sample consisting of residuals from both regimes. Precisely, the test statistic  $d_n$  is derived with the standard Gaussian cdf, the sample  $(y_1^1, y_2^1, \dots, y_{T-1}^1, y_1^2, y_2^2, \dots, y_{T-1}^2)$ , where  $(y_1^i, y_2^i, \dots, y_{T-1}^i)$  are the  $i$ th regime residuals, and the corresponding weights  $(w_1^1, w_2^1, \dots, w_{T-1}^1, w_1^2, w_2^2, \dots, w_{T-1}^2)$ , where  $w_t^i = \tilde{P}(R_t = i)$ .

Now, assume that the sample  $(x_1, x_2, \dots, x_T)$  is driven by the MRS model with independent regimes (specification I). 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:

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

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

$$g(X_{t,1}, \mathbf{x}_{t-1}) = \frac{X_{t,1} - \alpha - (1 - \beta)E(X_{t-1,1}|\mathbf{x}_{t-1})}{\sqrt{(1 - \beta)^2 \text{Var}(X_{t-1,1}|\mathbf{x}_{t-1}) + \sigma^2}}, \quad (17)$$

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

$$\begin{aligned} E(X_{t,1}|\mathbf{x}_t) &= P(R_t = 1|\mathbf{x}_t)x_t + \\ &\quad + P(R_t \neq 1|\mathbf{x}_t)[\alpha + (1 - \beta)E(x_{t-1,1}|\mathbf{x}_{t-1})], \end{aligned} \quad (18)$$

$$\begin{aligned} E(X_{t,1}^2|\mathbf{x}_t) &= P(R_t = 1|\mathbf{x}_t)x_t^2 + \\ &\quad + P(R_t \neq 1|\mathbf{x}_t)[\alpha^2 + 2\alpha(1 - \beta)E(X_{t-1,1}|\mathbf{x}_{t-1}) + \\ &\quad + (1 - \beta)^2 E(X_{t-1,1}^2|\mathbf{x}_{t-1}) + \sigma^2]. \end{aligned} \quad (19)$$

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 if  $k$  is such a number that  $P(R_{t-1} = 1|\mathbf{x}_{t-1}) = \dots = P(R_{t-k+1} = 1|\mathbf{x}_{t-k+1}) = 0$  and  $P(R_{t-k} = 1|\mathbf{x}_{t-k}) = 1$ , then  $g$  leads to transformation (7), i.e.  $g(X_{t,1}, \mathbf{x}_{t-1}) = h(X_{t,1}, x_{t-k,1}, k)$ .

Now, to test the fit of the mean-reverting regime, it is enough to calculate  $d_n$  according to formula (16) with the standard Gaussian cdf, the sample  $(y_1, y_2, \dots, y_{T-1})$ , where  $y_t = g(x_{t,1}, \mathbf{x}_{t-1})$  and the weights  $w_t = \tilde{P}(R_t = 1)$ . Observe that, the observations from the second regime are i.i.d. by definition, so the testing procedure is straightforward. The test statistic is derived with the  $F_2$  cdf, the sample of non-transformed observations  $(x_1, x_2, \dots, x_T)$  and the corresponding weights  $w_t = \tilde{P}(R_t = 2)$ . Again, to verify the goodness-of-fit of the whole model the subsamples corresponding to each of the regimes are combined. If  $H_0$  is true, then the resulting distribution should be a mixture of Gaussian and  $F_2$  laws. The test statistic  $d_n$  is, hence, calculated as in (16) with the mixture cdf (12) and the sample  $(y_1^1, y_2^1, \dots, y_{T-1}^1, y_1^2, y_2^2, \dots, y_{T-1}^2)$ , where  $(y_1^1, y_2^1, \dots, y_{T-1}^1)$  are the residuals, i.e.  $y_t^1 = g(x_{t,1}, \mathbf{x}_{t-1})$ , while  $(y_1^2, y_2^2, \dots, y_{T-1}^2)$  are the non-transformed observed variables, i.e.  $y_t^2 = x_t$ . The corresponding weights are then  $(w_1^1, w_2^1, \dots, w_{T-1}^1, w_1^2, w_2^2)$ , where  $w_t^i = \tilde{P}(R_t = i)$ .

Note, that like 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. Since the state process is latent, the weights  $w_t$  have to be estimated from the dataset. However, Lemma 1 is valid for known weights. 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 formula (16), higher or equal to the test statistic value obtained from the dataset.

#### 4 Simulations

In this Section we check the performance of the procedures introduced in Section 3.2. To this end, we generate 10000 trajectories of 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 random variables (Sim #1; with parameters  $\alpha_2$  and  $\sigma_2^2$ , 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  $\log(X) \sim N(\alpha_2, \sigma_2^2)$ , 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. Note, that the regimes of MRS models are not directly observable and, hence, the standard edf approach cannot be used.

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. 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 (19). 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%. This simple example clearly shows that in case of MRS models the wedf approach is more reliable.

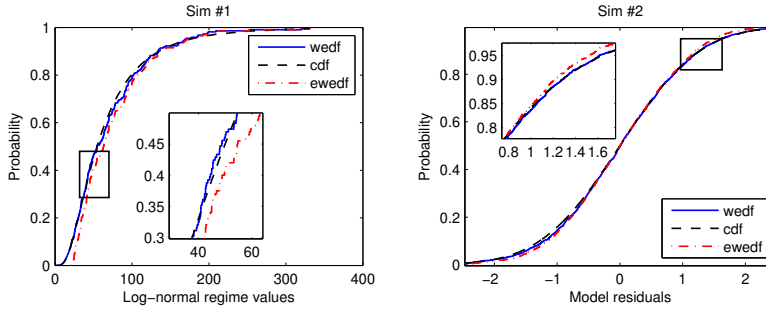
Further evidence is provided in Figure 1 where we illustrate the different types of empirical distribution functions. The wedf and ewedf functions are compared with the true edf. Note, that the edf can be calculated only when the simulated state process is known. Naturally, when dealing with real data,

**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	
	$\alpha_1$	$\beta_1$	$\sigma_1^2$	$\alpha_2$	$\beta_2$	$\sigma_2^2$	$p_{11}$	$p_{22}$
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

**Table 2** Percentage of rejected hypotheses  $H_0$  at the 5% significance level calculated from 10000 simulated trajectories of the models defined in Table 1. 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).

Regime	ewedf			wedf		
	First	Second	Model	First	Second	Model
Sim #1	0.0615	0.8646	0.0538	0.0538	0.0524	0.0572
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.

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 percentage given in Table 2.

Next, we compare the test results of the wedf approach based on the K-S test tables and Monte Carlo simulations. We calculate the  $p$ -values for sample

**Table 3** Means and standard deviations of the differences between  $p$ -values obtained by the wedf approach utilizing K-S test tables and Monte Carlo (MC) simulations. The calculations are based on 1000 simulated trajectories of the MRS model (for parameter details see Table 1). The number of Monte Carlo repetitions is equal to 1000.

		Regime		
		First	Second	Model
Sim #1	mean	0.0012	0.0018	-0.0162
	standard deviation	0.0134	0.0318	0.0147
Sim #2	mean	0.0015	0.0026	0.0130
	standard deviation	0.0165	0.0176	0.0137

MRS model trajectories Sim #1 and Sim #2 (for parameter details see Table 1). The number of repetitions in the Monte Carlo approach is 1000. The obtained results are given in Table 3. Observe that the means of the differences between  $p$ -values calculated by the two methods are close to zero and the standard deviations do not exceed 0.032. Hence, errors resulting from estimation of weights are quite low. However, if a  $p$ -value obtained from the K-S test tables is close to the significance level, the Monte Carlo approach should be used.

The simulation results presented so far were obtained with an assumption that the 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 recommended by Ross (2002), we use Monte Carlo simulations. We simulate 500 sample trajectories of the MRS model Sim #1 or Sim #2, estimate the model parameters and apply goodness-of-fit tests. We report the rejection percentage of the K-S test based on the Monte Carlo approach with 500 repetitions, as well as, the K-S test tables. For the Monte Carlo approach the procedure is as follows:

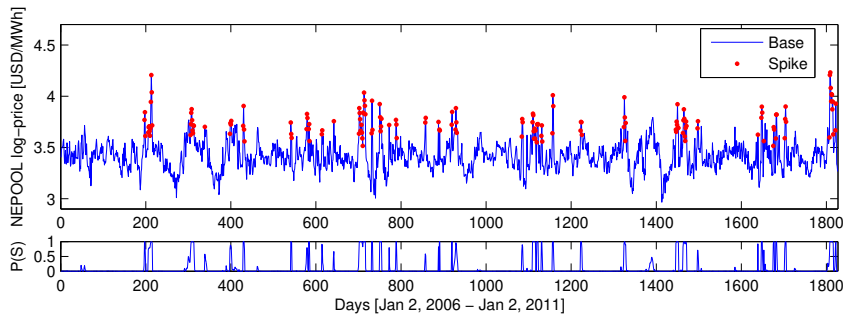
- estimate the parameter vector  $\hat{\theta}$  and calculate the test statistic  $d_n$  according to formula (11),
- simulate  $N$  trajectories with  $\hat{\theta}$ ,
- for each trajectory estimate the new parameter vector  $\hat{\theta}_i$  and calculate the new test statistic  $d_n^i$ ,
- calculate  $p$ -value as the proportion of simulated trajectories with the test statistic values higher or equal to  $d_n$ , i.e.  $\frac{1}{N} \#\{i : d_n^i \geq d_n\}$ .

The results obtained for the 5% significance level are given in Table 4.

Looking at the test results based on the K-S test tables, for the ewdf approach the rejection percentage is again too high. On the other hand, for the wedf approach the  $p$ -values are overestimated, what results in the rejection percentage 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 wrongly not reject a false  $H_0$  hypothesis. This is not the case for the wedf approach with Monte Carlo simulations as the obtained rejection percentage is close to the 5% significance level. This example clearly shows that the results of the

**Table 4** Percentage of rejected hypotheses  $H_0$  at the 5% significance level calculated from 500 simulated trajectories of the models defined in Table 1 with parameters estimated from the 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 500 repetitions.

		ewedf			wedf		
	Regime	First	Second	Model	First	Second	Model
Sim #1	K-S test tables	0.0280	0.5680	0.0280	0.0320	0	0.0280
	MC simulation	0.0520	0	0.0560	0.0560	0.0340	0.0560
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



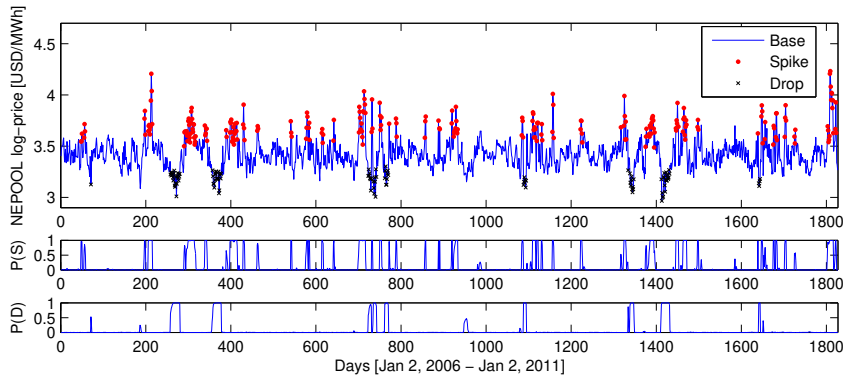
**Fig. 2** Calibration results for the 2-regime MRS model with a mean reverting base regime and independent log-normally distributed ‘spikes’ fitted to NEPOOL log-prices. Observations with  $\tilde{P}(R_t = 2) > 0.5$ , i.e. the ‘spikes’, are denoted by dots. The lower panel displays the probability 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

wedf test based on the K-S test tables can only be used if the test 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.

## 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, 2003; 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 abrupt short-lived price changes called spikes. To cope with the seasonality we use the standard time series decomposition approach and let the electricity spot price  $P_t$  be repre-



**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  $\hat{P}(R_t = 2) > 0.5$  or  $\hat{P}(R_t = 3) > 0.5$ , i.e. the ‘spikes’ or ‘drops’, are denoted by dots or ‘x’ in the upper panel. The lower panels display the probability  $\hat{P}(R_t = 2)$  or  $\hat{P}(R_t = 3)$  of being in the ‘spike’ or ‘drop’ regime, respectively.

sented 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, 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 ( $R_t = 2$ )
- 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  $\log(X - q)$  ( $\log(q - X)$ ) has the Gaussian distribution. Note that  $q$  can be chosen arbitrarily, however, here 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 natural gas, 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 (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 minimum of the new process  $X_t$  is the same as the minimum of  $P_t$ . The



**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		
	$\alpha$	$\beta$	$\sigma^2$	$\mu_2$	$s_2^2$	$\mu_3$	$s_3^2$	$p_{11}$	$p_{22}$	$p_{33}$
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

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

Regime	ewdf				wdf			
	Base	Spike	Drop	Model	Base	Spike	Drop	Model
<i>2-regime model</i>								
K-S test tables	<b>0.21</b>	<b>0.27</b>	-	<b>0.12</b>	<b>0.08</b>	<b>0.93</b>	-	<b>0.10</b>
MC simulations	0.01	<b>0.07</b>	-	0.01	0.00	<b>0.57</b>	-	0.00
<i>3-regime model</i>								
K-S test tables	<b>0.56</b>	<b>0.25</b>	<b>0.98</b>	<b>0.68</b>	<b>0.38</b>	<b>0.71</b>	<b>0.92</b>	<b>0.51</b>
MC simulations	<b>0.19</b>	<b>0.06</b>	<b>0.81</b>	<b>0.20</b>	<b>0.31</b>	<b>0.27</b>	<b>0.49</b>	<b>0.25</b>

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 ewdf, as well as, the wdf approach is performed. Moreover, since the standard approach based on the K-S test tables might produce overestimated  $p$ -values, the Monte Carlo results are also provided. The obtained  $p$ -values are reported in Table 6. For the 2-regime model the  $p$ -values obtained from the 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. Observe that 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 wdf-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 carried on in this direction.

## 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 and the seminars at Macquarie University, University of Sydney and University of Verona. J.J. acknowledges partial financial support from the European Union within the European Social Fund.

## Appendix

*Proof (Lemma 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  $D = \mathbb{R} \setminus \{x : F(x) = 0 \vee F(x) = 1\}$ . Therefore in the following we will limit ourselves to the case  $0 < F(x) < 1$ .

Second, observe that the distribution of  $\sup_{x \in \mathbb{R}} |F_n(x) - F(x)|$  does not depend on  $F$ . Indeed, since  $U_t = F(Y_t)$  has the uniform distribution,

$$P\left(\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \leq u\right) = P\left(\sup_{x \in \mathbb{R}} \left|\sum_{t=1}^n \frac{w_t \mathbb{I}_{\{U_t < y\}}}{\sum_{t=1}^n w_t} - y\right| \leq u\right),$$

where  $y = F(x)$ .

Next, note that the sequence of random variables  $Z_t = w_t \mathbb{I}_{\{Y_t < x\}}$  satisfies the Lindeberg condition (Billingsley, 1986, p.369). Let  $S_n^2 = \sum_{t=1}^n \text{Var}(Z_t)$  and  $\mu_t = E(Z_t)$ , where

$Var(Z_t) = F(x)[1 - F(x)]w_t^2$  and  $E(Z_t) = w_t F(x)$ . The Lindeberg condition yields

$$\begin{aligned}
& \frac{1}{S_n^2} \sum_{t=1}^n \int_{\{|Z_t - \mu_t| > \delta S_n\}} (Z_t - \mu_t)^2 dP = \\
& = \frac{1}{S_n^2} \sum_{t=1}^n w_t^2 \int_{\{w_t |\mathbb{I}_{\{Y_t < x\}} - F(x)| > \delta S_n\}} [\mathbb{I}_{\{Y_t < x\}} - F(x)]^2 dP \leq \\
& \leq \frac{1}{S_n^2} \sum_{t=1}^n w_t^2 P(w_t |\mathbb{I}_{\{Y_t < x\}} - F(x)| > \delta S_n) \max\{F(x)^2, [1 - F(x)]^2\} \leq \\
& \leq \frac{1}{\sum_{t=1}^n w_t^2 [1 - F(x)] F(x)} \sum_{t=1}^n w_t^2 P(w_t |\mathbb{I}_{\{Y_t < x\}} - F(x)| > \delta S_n) \leq \\
& \leq \frac{\max_{1 \leq t \leq n} P(w_t |\mathbb{I}_{\{Y_t < x\}} - F(x)| > \delta S_n)}{[1 - F(x)] F(x)}.
\end{aligned}$$

Since  $w_t |\mathbb{I}_{\{Y_t < x\}} - F(x)| \leq M$  and the fact that  $\lim_{n \rightarrow \infty} \sum_{t=1}^n w_t^2 = \infty$  implies  $\lim_{n \rightarrow \infty} S_n = \infty$  we have

$$\exists n_0 \forall n > n_0 \forall 1 \leq t \leq n \quad P(w_t |\mathbb{I}_{\{Z_t < x\}} - F(x)| > \delta S_n) = 0. \quad (20)$$

Therefore, the Lindeberg condition is satisfied:

$$\lim_{n \rightarrow \infty} \frac{1}{S_n^2} \sum_{t=1}^n \int_{\{|Z_t - \mu_t| > \delta S_n\}} (Z_t - \mu_t)^2 dP = 0. \quad (21)$$

This ensures that the Central Limit Theorem holds for  $Z_1, Z_2, \dots$  and

$$\frac{\sum_{t=1}^n w_t \mathbb{I}_{\{Y_t < x\}} - \sum_{t=1}^n w_t F(x)}{\sqrt{F(x)[1 - F(x)] \sum_{t=1}^n w_t^2}} \xrightarrow{d} N(0, 1). \quad (22)$$

The latter is equivalent to

$$\frac{\sum_{t=1}^n w_t}{\sqrt{\sum_{t=1}^n w_t^2}} [F_n(x) - F(x)] \xrightarrow{d} N(0, F(x)[1 - F(x)]). \quad (23)$$

Recall that the Kolmogorov-Smirnov distribution  $KS$  is a distribution of  $\sup_{0 \leq y \leq 1} |B(y)|$ , where  $B(y)$  is a Brownian bridge, i.e.  $B(y) \sim N(0, y(1 - y))$ , see e.g. Lehmann and Romano (2005), p. 585. Therefore, putting  $y = F(x)$  and taking the supremum, we obtain that

$$\frac{\sum_{t=1}^n w_t}{\sqrt{\sum_{t=1}^n w_t^2}} \sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \xrightarrow{d} KS. \quad (24)$$

## 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. (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.
- Čížek 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.
- 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. (2003) *Energy and Power Risk Management*. Wiley, Hoboken, NJ.
- Garcia, R. (1998) Asymptotic null distribution of the likelihood ratio test in Markov switching models. *International Economic Review* 39, 763-788.
- 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. (2011) Efficient estimation of Markov regime-switching models: An application to electricity spot prices. *AStA – Advances in Statistical Analysis*, *submitted*. Earlier working paper version available at MPRA: <http://mpra.ub.uni-muenchen.de/26628/>.
- 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.
- 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.
- 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.