An empirical comparison of alternate regime-switching models or electricity spot prices

Janczura, Joanna and Weron, Rafal

6 February 2010

An empirical comparison of alternate regime-switching models for electricity spot prices

Joanna Janczura

Hugo Steinhaus Center, Institute of Mathematics and Computer Science, Wrocław University of Technology, 50-370 Wrocław, Poland

Rafał Weron

Institute of Organization and Management, Wrocław University of Technology, 50-370 Wrocław, Poland

---

Abstract

One of the most profound features of electricity spot prices are the price spikes. Markov regime-switching (MRS) models seem to be a natural candidate for modeling this spiky behavior. However, in the studies published so far, the goodness-of-fit of the proposed models has not been a major focus. While most of the models were elegant, their fit to empirical data has either been not examined thoroughly or the signs of a bad fit ignored. With this paper we want to fill the gap. We calibrate and test a range of MRS models in an attempt to find parsimonious specifications that not only address the main characteristics of electricity prices but are statistically sound as well. We find that the most universal and robust structure is that of an independent spike 3-regime model with heteroscedastic diffusion-type base regime dynamics and shifted spike regime distributions.

Keywords: Electricity spot price, Spikes, Markov regime-switching, Heteroscedasticity.

---

1. Introduction

The valuation of electricity contracts is not a trivial task. If the model is too complex the computational burden will prevent its on-line use in trading departments. On the other hand, if the price process chosen is inappropriate to capture the main characteristics of electricity prices, the results from the model are unlikely to be reliable.

The uniqueness of electricity as a commodity prevents us from simply using models developed for the financial or other commodity markets. Electricity cannot be stored economically and requires immediate delivery, while end-user demand shows high variability and strong weather and business cycle dependence. Effects like power plant outages or transmission grid (un)reliability add complexity and randomness. The resulting spot price series exhibit strong seasonality on the annual, weekly and daily level, as well as, mean reversion, very high volatility and abrupt, short-lived and generally unanticipated extreme price changes known as spikes or jumps.

Despite numerous attempts (for reviews see e.g. Benth et al., 2008; Bunn, 2004; Kaminski, 2004; Weron, 2006), the need for realistic models of price dynamics capturing the unique characteristics of electricity and adequate derivatives pricing techniques still has not been fully satisfied. It is the aim of this paper to suggest parsimonious models for electricity spot price dynamics that not only address the main characteristics of electricity prices but are statistically sound as well. The parsimony is a prerequisite of derivatives pricing, especially simulation techniques, where the numerical burden can be substantial. The statistical adequacy, on the other hand, is a requirement in any modeling task.

We focus on Markov regime-switching (MRS) models, which seem to be a natural candidate for modeling the spiky, non-linear behavior of electricity spot prices. In a way, they offer the best of the two worlds; they are a trade-off

---

Email addresses: joanna.janczura@pwr.wroc.pl (Joanna Janczura), rafal.weron@pwr.wroc.pl (Rafał Weron)

between model parsimony and adequacy to capture the unique characteristics of power prices. In contrast to threshold type regime-switching models (like TAR, STAR, SETAR), in MRS models the regimes are only latent. Consequently, MRS models do not require an upfront specification of the threshold variable and level and, hence, are less prone to modeling risk.

Yet, despite this latency, there are still enough 'free parameters' in MRS models to make the calibration procedure a tough exercise, even for experienced professionals. Firstly, the number of regimes has to be agreed upon. In almost all published studies only 2-regime models were considered, most likely due to a lower computational burden. Apart from a base regime, a spike (or excited) regime was introduced for modeling the extreme price behavior. However, there is no fundamental reason for not considering 3- or multi-regime specifications. In fact, analyzing UK half-hourly electricity spot prices Karakatsani and Bunn (2008) identified three regimes, with the third regime capturing the most extreme prices. The estimated switching pattern suggested a two-stage spike reversal to normal prices. Also, for many of the very low prices a technical (or fundamental) event underlying the non-standard behavior can be identified, justifying existence of a separate 'down-spike' or 'drop' regime.

Secondly, the stochastic processes defining the price dynamics in each of the regimes have to be selected. The base regime is typically of a mean-reverting diffusion-type, however, for the spike regime(s) a number of specifications have been suggested, ranging from mean reverting diffusions to heavy tailed random variables.

Finally, the dependence between the regimes has to be decided upon. Dependent regimes with the same random noise process in all regimes (but different parameters) lead to computationally simpler models. On the other hand, independent regimes allow for a greater flexibility and seem to be a more natural choice for a process which, from time to time, radically changes its dynamics.

This paper is intended as a guide to MRS models for spot electricity prices. In Section 2 we review the first and second generation models and discuss their pros and cons. In Section 3 we present the datasets and explain the deseasonalization procedures. Next, in Section 4 we discuss the estimation issues and goodness-of-fit testing. The more technical issues are discussed in Appendix A and Appendix B, respectively. In Section 5 we calibrate various MRS models to deseasonalized prices and log-prices, evaluate their goodness-of-fit and select the optimal model structure. Finally, in Section 6 we conclude.

2. Overview of MRS models for spot electricity prices

The underlying idea behind the Markov regime-switching (MRS) scheme is to model the observed stochastic behavior of a specific time series by two (or more) separate phases or regimes with different underlying processes. In other words, the parameters of the underlying process may change for a certain period of time and then fall back to their original structure. The switching mechanism between the states is Markovian and is assumed to be governed by an unobserved (latent) random variable. The underlying processes, though, do not have to be Markovian, but are often assumed to be independent from each other.

In the simplest case of a 2-regime model, the spot price can be assumed to display either normal or very high prices (or spikes) at each point in time, depending on the regime $R_t = b$ or $R_t = s$. Consequently, we have a probability law that governs the transition from one state to another. The transition matrix $\mathbf{P}$ contains the probabilities p_{ij} of switching from regime i at time t to regime j at time $t+1$, for $i, j = \{b, s\}$:

$$\mathbf{P} = (p_{ij}) = \begin{pmatrix} p_{bb} & p_{bs} \\ p_{sb} & p_{ss} \end{pmatrix} = \begin{pmatrix} 1-p_{bs} & p_{bs} \\ p_{sb} & 1-p_{sb} \end{pmatrix}. \tag{1}$$

In the more general case of a 3-regime model, the transition matrix $\mathbf{P}$ takes the form:

$$\mathbf{P} = (p_{ij}) = \begin{pmatrix} p_{bb} & p_{bs} & p_{bd} \\ p_{sb} & p_{ss} & p_{sd} \\ p_{db} & p_{ds} & p_{dd} \end{pmatrix}, \text{ with } p_{ii} = 1 - \sum_{j \neq i} p_{ij}. \tag{2}$$

Because of the Markov property the current state R_t at time t of a Markov chain depends on the past only through the most recent value R_{t-1}. Consequently the probability of being in state j at time $t+m$ starting from state 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 denotes the ith column of the identity matrix.

2.1. First generation models

To our best knowledge, the MRS models were first applied to electricity prices in Deng (1998) and Ethier and Mount (1998). Deng (1998) considered three models of spot price dynamics in a derivatives pricing context (formal parameter estimation was not performed). These included a 2-state regime-switching specification for the log-prices in which the base regime was driven by an autoregressive process of order one, i.e. AR(1), and the spike regime by the same AR(1) process (i.e. with the same parameters) shifted by an exponentially distributed random variable (spike/jump size). In a parallel paper Ethier and Mount (1998) proposed a 2-state model with mean-reverting AR(1) processes for the log-prices in both regimes. The processes shared the same set of random innovations – only the parameters of the two AR(1) processes varied between the regimes. Strong empirical support for the existence of different means and variances in the two regimes was found for data from four U.S. and Australian markets. A similar model was later used by Heydari and Siddiqui (2010) in a gas-fired power plant valuation problem in the UK market, but was found inferior to mean-reverting stochastic volatility models over long-term forecast periods.

Huisman and Mahieu (2003) proposed a regime-switching model with three possible states in which the initial jump regime was immediately followed by the reversing regime and then moved back to the base regime, i.e. with $p_{sd} = p_{db} = 1$ in (2). Andreasen and Dahlgren (2006) studied a similar MRS model – in their model the spot price could jump up or down if it was in the neutral state (base regime) and it could only jump back to its neutral state if it was in one of the jump regimes. Consequently, these models did not allow for consecutive high prices (in fact of log-prices) and, hence, did not offer any obvious advantage over jump-diffusion models.

This restriction was relaxed by Huisman and de Jong (2003), who proposed a simple independent spike (IS) 2-regime model for deseasonalized log-prices. The base regime was modeled by a mean-reverting AR(1) process and the spike regime by a normal distributed random variable whose mean and variance were higher than those of the base regime process. The third regime was not needed to pull prices back to stable levels, because the prices were independent from each other in the two regimes. Kosater and Mosler (2006) reported that for EEX 2000-2004 data the IS 2-regime model outperformed the model of Ethier and Mount (1998) in terms of short/medium-term forecasts (up to 1 month), while the opposite was true for long-term forecasts. They were able to improve the forecasts if the spike regimes in these models had two sets of parameters – one for business days and one for holidays.

The IS 2-regime model was further investigated by Weron et al. (2004), who introduced log-normally distributed spikes. In a follow up paper Bierbrauer et al. (2004) tested whether using Pareto distributed spikes would lead to better models for deseasonalized Nord Pool log-prices from the period 1997-2000. This was not the case, as the Pareto distribution seemed to overestimate the spike sizes. The best model (in terms of the likelihood) turned out to be the one with log-normal spikes, closely followed by the model of Huisman and de Jong (2003), i.e. with Gaussian spikes. The 3-regime model of Huisman and Mahieu (2003) came in last.

De Jong (2006) proposed another modification of the basic IS 2-regime model with autoregressive, Poisson driven spike regime dynamics (a similar model was later considered by Mari, 2008). Using 2001-2006 data from eight European and U.S. power markets he compared it to several spot price models. On average the new model yielded the best fit (again in terms of the likelihood), but the IS 2-regime model with Gaussian spikes was nearly as good. Like in the previously mentioned study, the 3-regime model of Huisman and Mahieu (2003) yielded a significantly worse fit, comparable to that of a mean-reverting jump diffusion (MRJD) and slightly worse than that of the threshold model of Geman and Roncoroni (2006). In a related study Bierbrauer et al. (2007) introduced an IS 2-regime model with exponentially distributed spikes. For 2000-2003 EEX market data they found it inferior to MRS models with normal or log-normal spikes, but much better than the non-linear mean-reverting model of Barlow (2002) or the MRJD model of Kluge (2006) with two different mean-reversion rates for the normal and the jump parts.

2.2. Second generation models

The above mentioned first generation models have two common features. Firstly, not the prices X_t themselves, but rather log-prices $Y_t = \log(X_t)$ are considered. Secondly, the base regime (and in some cases the spike regime as well) is driven by a mean-reverting diffusion process of the form:

$$dY_t = (\alpha - \beta Y_t)dt + \sigma dW_t = \beta\left(\frac{\alpha}{\beta} - Y_t\right)dt + \sigma dW_t, \tag{4}$$

where W_t is Brownian motion (i.e. a Wiener process), $\frac{\alpha}{\beta}$ is the long term mean reversion level, β is the speed of mean reversion and σ is the volatility. In the fixed income literature this popular process is known as the Vasicek (1977) model, in mathematics as the (generalized) Ornstein-Uhlenbeck process (Janicki and Weron, 1994), while in signal processing – when discretized – as an AR(1) process or an autoregressive time series of order one (Brockwell and Davis, 2002). But is this the right choice for the base regime dynamics? Should we model prices or log-prices? What about the regime-switching mechanism – perhaps we could use fundamental information to improve the fit?

The second generation models tried to address some of the deficiencies of the first generation models. They can be grouped into two categories: (i) those that use fundamental information (system constraints, weather variables) to better model regime-switching and (ii) those that propose statistical refinements (to ultimately improve the goodness-of-fit).

In the first group, Mount et al. (2006) proposed a 2-regime model with two AR(1) regimes for log-prices and transition probabilities dependent on the reserve margin. Using PJM data from 1999-2000 they showed that the estimated switching probability from the base (low) to the spike (high) regime predicts price spikes well if the reserve margin is measured accurately. This is in line with the qualitative findings of Kanamura and Ōhashi (2008), who showed that the transition probabilities cannot be constant, but depend on the current demand level relative to the supply capacity, the deterministic trend of demand change, and the trend generated by the deviation of temporary demand fluctuation from its long-term mean.

Anderson and Davison (2008) went one step further and explicitly modeled power plant failure and repair in a hybrid 2-regime model, which included sub-models for forced outages, planned outages and load. Using PJM data from the years 2000-2003 they concluded that the model performed better when modeling the seasons where the planned outages were known. In a non-MRS context, studying the 2003-2006 data for the England and Wales market, Cartea et al. (2009) reached similar conclusions. They showed that the incorporation of forward looking information on capacity constraints significantly improved the modeling of spikes, both timing and magnitude.

In a complementary study to Mount et al. (2006), Huisman (2008) noted that the availability (to every market participant) of the reserve margin data is limited. Hence, he proposed to use temperature as a proxy. Interpreting the results from three MRS models fitted to the Dutch APX log-prices from the period 2003-2008, Huisman showed that the probability of spike occurrence increases when temperature deviates substantially from mean temperature levels. However, in general, temperature does not provide as much information as the reserve margin.

In the second group of models the focus has been on refining the statistical tools used for describing the dynamics of electricity prices. Examples range from non-orthodox approaches originating in physics to more 'traditional' econometric ones. For instance, Lucheroni (2009) suggested to model a seasonally and irregularly peaking price dynamics using a FitzHugh–Nagumo system of coupled nonlinear stochastic differential equations. The rationale behind this approach stems from the fact that second order dynamics is obviously richer than first order (like in autoregressions). It can sustain oscillations even without periodic driving, and, when driven, its behavior can be very complex.

The more 'traditional' econometric refinements have been aimed at two common flaws of the first generation models: negative 'expected spike sizes' and regime misclassification. While most of the mentioned above models were elegant, their fit to empirical data has either been not examined thoroughly or the signs of a bad fit ignored.

Some authors reported that the 'expected spike sizes' ($\equiv E(Y_{t,spike}) - E(Y_{t,base})$) were negative (see e.g. De Jong, 2006; Bierbrauer et al., 2007). However, these findings were not considered as evidence for model misspecification. Further, regime classification was not checked but, as Weron (2009) showed, for log-price models the calibration scheme generally assigned all extreme prices to the spike regime, no matter whether they truly were spikes or only sudden drops, see Figure 1. This 'regime misclassification' leads to a discrepancy between the often bimodal empirical distribution of (log-)prices classified as spikes and the unimodal theoretical spike distribution assumed in practically all MRS models, see Figure 2. Based on these findings Weron (2009) conjectured that – contrary to the common belief that electricity price models 'should be built on log-prices' – modeling the prices themselves may be more beneficial and methodologically sound in some cases.

In a follow up paper Janczura and Weron (2009) tested IS 2-regime models for electricity spot prices with Vasicek (see eqn. (4)) and CIR-type (see Cox et al., 1985) dynamics for the base regime and median-shifted spike regime distributions. For EEX market data from the period 2001-2009 they found that models with shifted spike regime distributions (which assign zero probability to prices below the median price) led to more realistic descriptions of electricity spot prices and that by introducing CIR-type heteroscedasticity in the base regime – in place of the standard mean-reverting, constant volatility dynamics – better spike identification was obtained.

We continue this line of research and in the following sections calibrate and test a range of MRS models in an

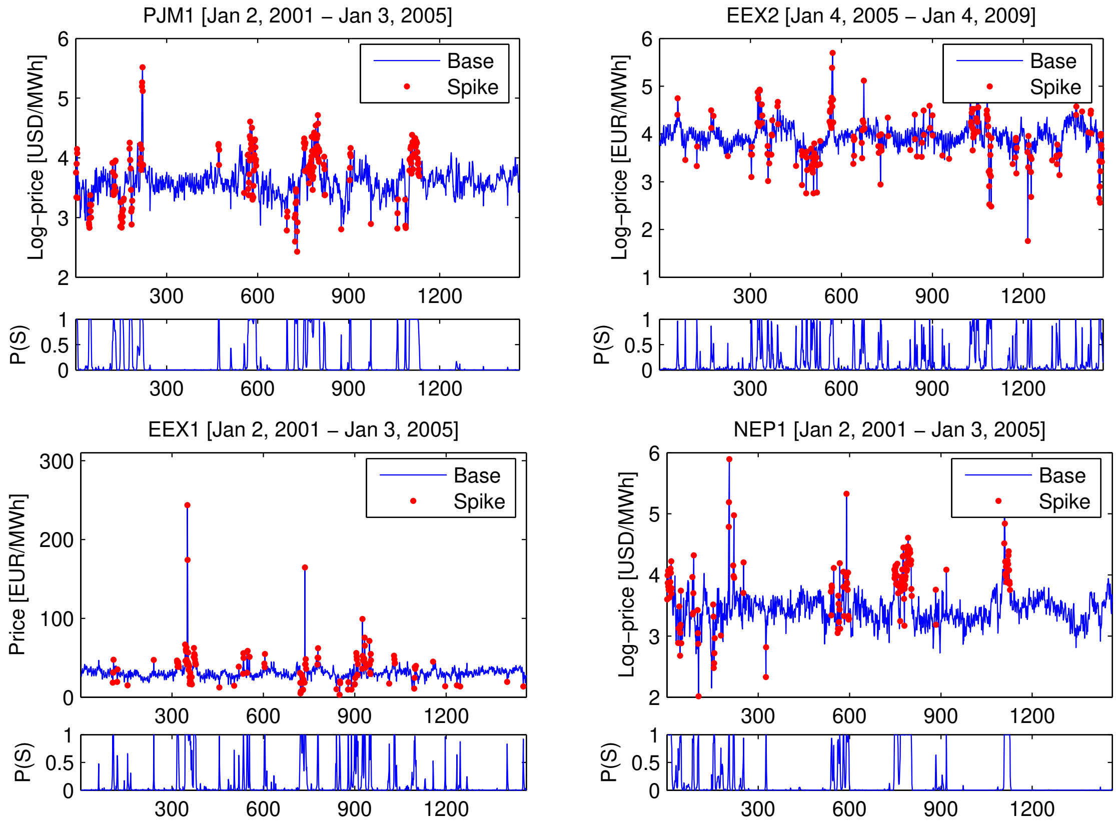

Figure 1: Sample calibration results for 2-regime models with Vasicek, i.e. AR(1), base regime dynamics and alternative spike regimes fitted to deseasonalized prices or log-prices from three major power markets. *Top left*: An independent spike (IS) model with normal spikes fitted to PJM log-prices. *Top right*: The Ethier and Mount (1998) model with AR(1) spike regime fitted to EEX log-prices. *Bottom left and right*: An IS model with lognormal spikes fitted to EEX prices and NEPOOL log-prices, respectively. The corresponding lower panels display the probability $P(S) \equiv P(R_t = s)$ of being in the spike regime. The prices or log-prices classified as spikes, i.e. with $P(S) > 0.5$, are additionally denoted by dots in the upper panels. For descriptions of the datasets see Section 3 and Figure 3; for model details see Section 5.1.

attempt to find parsimonious specifications that not only address the main characteristics of electricity prices but are statistically sound as well.

It is interesting to note that, to our best knowledge, Janczura and Weron (2009) was the first paper in the 'MRS electricity spot price modeling literature' where actually hypothesis testing (based on the Kolmogorov-Smirnov test) was performed. We further improve the testing methodology (see Appendix B) and use it to identify correctly specified MRS models.

3. Data preprocessing

3.1. The datasets

In this study we use mean daily (baseload) day-ahead spot prices from three major power markets: the European Energy Exchange (EEX; Germany), the PJM Interconnection (PJM; U.S.) and the New England Power Pool (NEPOOL; U.S.). For each market the sample totals 2926 daily observations (or 418 full weeks) and covers the 8-year period January 2, 2001 – January 4, 2009. To see how the presented methods perform under different market conditions each dataset is split into two subsamples of 1463 daily observations (209 weeks each): January 2, 2001 – January 3, 2005 (EEX1, PJM1 and NEP1) and January 4, 2005 – January 4, 2009 (EEX2, PJM2 and NEP2), see Figure 3. Note, that starting in late 2004 the spot prices exhibit an upward trend and higher volatility, largely due to a combination of higher fuel prices and the introduction of CO_2 emission costs in Europe in January 2005 (Benz and Trück,

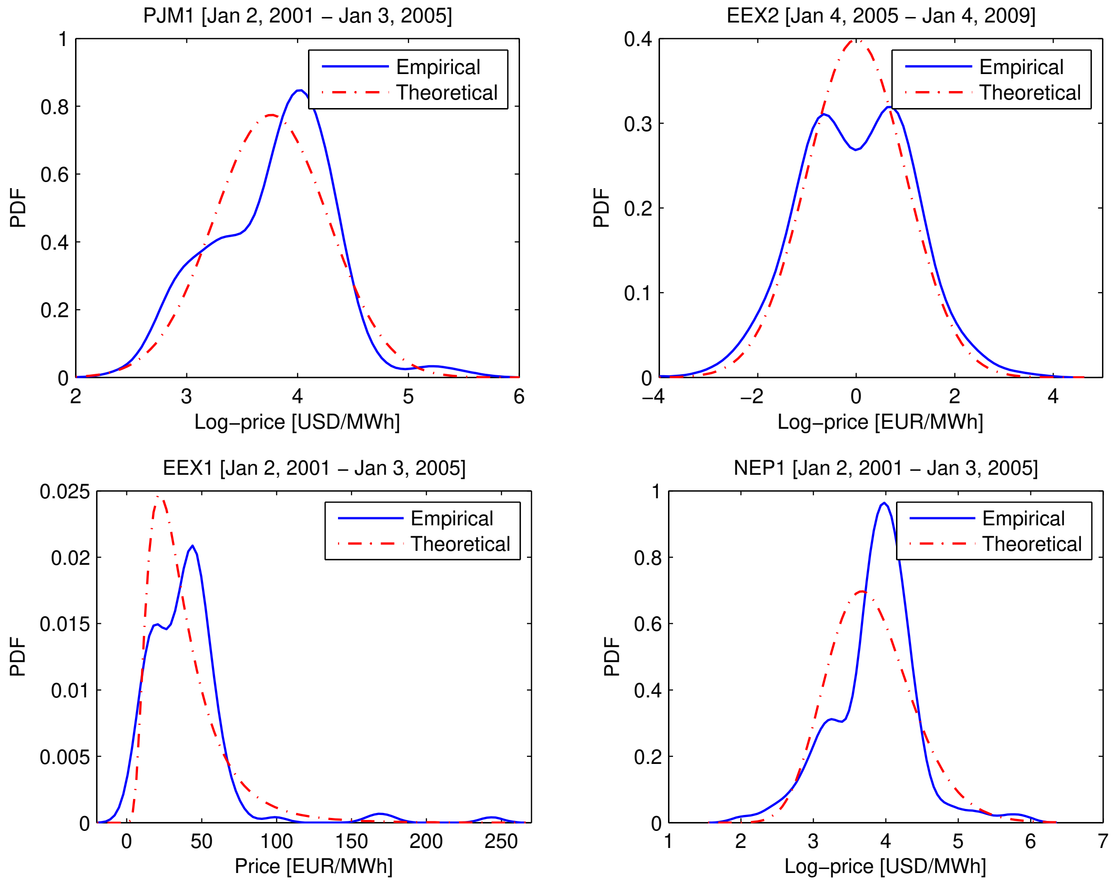

Figure 2: Comparison of empirical (sample) and theoretical (model implied) spike regime probability distribution functions in the 2-regime models. The models and datasets are the same as in Figure 1. Note, that for the Ethier and Mount (1998) model the distributions of the noise in the AR(1) process driving the spike regime are plotted (*top right*).

2006; Weron, 2009). The influence of fuel prices on electricity can be also seen in the last months of the second period. Due to the sub-prime crisis in the U.S., the global economy went into a recession and the demand for energy products has dramatically decreased. This effect can be seen first in the U.S. markets (PJM and NEPOOL) and a few months later in the European market (EEX).

3.2. *Deseasonalization*

The first crucial step in defining a model for electricity price dynamics consists of finding an appropriate description of the seasonal pattern. There are different suggestions in the literature for dealing with this task (Trück et al., 2007). Here we follow the ‘industry standard’ and represent the spot price P_t by a sum of two independent parts: a predictable (seasonal) component f_t and a stochastic component X_t, i.e. $P_t = f_t + X_t$. Further, we let f_t be composed of a weekly periodic part s_t and a long-term seasonal trend T_t, which represents both the changing climate/consumption conditions throughout the year and the long-term non-periodic structural changes.

As in Weron (2009) the deseasonalization is conducted in three steps. First, T_t is estimated from daily spot prices P_t using a wavelet filtering-smoothing technique (for details see Trück et al., 2007; Weron, 2006). Recall, that any function or signal (here: P_t) can be built up as a sequence of projections onto one father wavelet and a sequence of mother wavelets: $S_J + D_J + D_{J-1} + ... + D_1$, where 2^J is the maximum scale sustainable by the number of observations. At the coarsest scale the signal can be estimated by S_J. At a higher level of refinement the signal can be approximated by $S_{J-1} = S_J + D_J$. At each step, by adding a mother wavelet D_j of a lower scale $j = J-1, J-2, ...$, we obtain

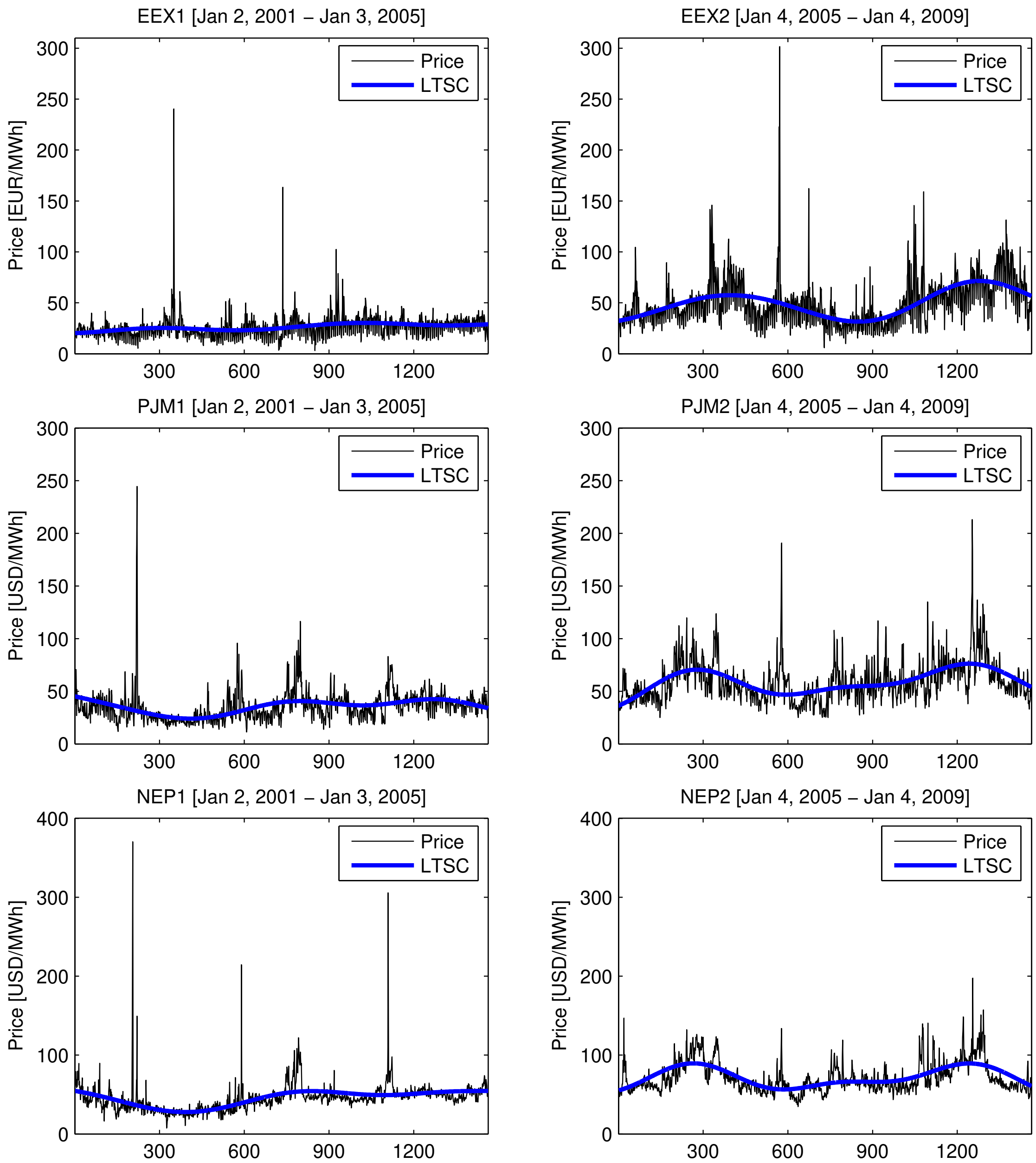

Figure 3: Mean daily spot prices and their long-term seasonal components (LTSC; thick blue lines) for EEX, PJM and NEPOOL power markets. The first four years (January 2, 2001 – January 3, 2005) of the study period are presented in the left panels and the latter four years (January 4, 2005 – January 4, 2009) in the right panels.

a better estimate of the original signal. Here we use the S_8 approximation, which roughly corresponds to annual ($2^8 = 256$ days) smoothing, see the thick blue lines in Figure 3. The price series without the long-term seasonal trend is obtained by subtracting the S_8 approximation from P_t. Next, the weekly periodicity s_t is removed by applying the moving average technique (see e.g. Brockwell and Davis, 2002; Weron, 2006) and subtracting the resulting 'mean' weekly pattern. Finally, the deseasonalized prices, i.e. $P_t - T_t - s_t$, are shifted so that the minimum of the new process is the same as the minimum of P_t (the latter alignment is required if log-prices are to be analyzed). The resulting deseasonalized time series X_t can be seen in Figure 6.

4. Statistical inference for MRS models

Calibration of MRS models is not straightforward since the regime is only latent and hence not directly observable. Here we follow Hamilton (1990), Kim (1994) and Janczura and Weron (2010) and apply a variant of the Expectation-Maximization (EM) algorithm. This is an iterative two-step procedure which starts with computing the conditional probabilities $P(R_t = j|x_1, ..., x_T; \theta)$ for the process being in regime j at time t, based on starting values $\hat{\theta}^{(0)}$ for the parameter vector θ of the underlying stochastic processes. These probabilities are referred to as 'smoothed inferences'. Then, in the second step, new and more exact ML estimates $\hat{\theta}$ for all model parameters are calculated using the smoothed inferences from step one. With each new vector $\hat{\theta}^{(n)}$ the next cycle of the algorithm is started in order to reevaluate the smoothed inferences. Every iteration of the EM algorithm generates new estimates $\hat{\theta}^{(n+1)}$, as well as, new estimates for the smoothed inferences. Each iteration cycle increases the log-likelihood function and the limit of this sequence of estimates reaches a (local) maximum of the log-likelihood function. For a detailed description of the estimation algorithm used in this paper, see Appendix A.

In order to evaluate the goodness-of-fit, we report basic descriptive statistics. These include the Inter-Quartile and the Inter-Decile Range, i.e. the difference between the third and the first quartiles (IQR) or ninth and first deciles (IDR). The quantile-based measures rather than the less robust to outliers moment-related statistics are used. Moreover, we report the p-values of the Kolmogorov-Smirnov goodness-of-fit test (K-S test) measuring the difference between the empirical (sample) and theoretical (model implied) marginal distributions. Since the K-S test cannot be applied directly to prices (or log-prices) as they are not i.i.d. in the considered models, following Janczura and Weron (2010) we transform the data using the smoothed inferences and obtain a mixture of i.i.d. samples resulting from the two (or three) regimes. The K-S test is then performed for the regimes, as well as, for the whole sample (for details see Appendix B).

5. Empirical results

In this section we calibrate various MRS models to deseasonalized prices and log-prices and evaluate their goodness-of-fit. We briefly recall the first generation models, then continue with more elaborate specifications. The ultimate objective is finding the optimal model structure – a parsimonious specification that not only addresses the main characteristics of electricity prices but is statistically sound as well.

5.1. First generation 2-regime models

The 2-regime specifications of the first generation models share a common property – the base regime dynamics is given by a mean reverting Vasicek stochastic differential equation:

$$dX_{t,b} = (\alpha - \beta X_{t,b})dt + \sigma_b dW_t, \tag{5}$$

where $\alpha, \beta, \sigma_b = \text{const.}$, W_t is Brownian motion and $X_{t,b}$ denotes the value of the base regime process at time t. Applying the Euler scheme to (5) we obtain the discrete time AR(1) process:

$$X_{t,b} = \alpha - (1 - \beta)X_{t-1,b} + \sigma_b \epsilon_t, \tag{6}$$

where ϵ_t has the standard Gaussian distribution.

In the model of Ethier and Mount (1998) the spike regime has the same dynamics, only with different parameters, say, α_s, β_s and σ_s. The noise term (ϵ_t) is shared by both regimes, which significantly simplifies the estimation process (see Appendix A). In the independent spike (IS) models the dynamics of the spike regime is described by a certain probability distribution. Huisman and de Jong (2003) used the Gaussian:

$$X_{t,s} \sim \mathrm{N}(\mu_s, \sigma_s^2), \tag{7}$$

while Bierbrauer et al. (2004) advocated the use of the lognormal:

$$\log(X_{t,s}) \sim \mathrm{N}(\mu_s, \sigma_s^2), \tag{8}$$

or the Pareto distribution:

$$X_{t,s} \sim F_{\mathrm{Pareto}}(\sigma_s, \mu_s) = 1 - \left(\frac{\mu_s}{x}\right)^{\sigma_s}, x > \mu_s. \tag{9}$$

The main problem with the first generation models is that they possess a critical flaw. Namely, they do not classify the regimes correctly. This flaw, in most cases, leads to negative 'expected spike sizes', i.e. $E(Y_{t,spike}) < E(Y_{t,base})$, and a significant discrepancy between the often bimodal empirical distribution of (log-)prices classified as spikes and the unimodal theoretical spike distribution, see Figures 1 and 2. To eliminate this flaw Weron (2009) proposed to build models not for log-prices, but prices themselves. Indeed, this approach alleviates the unwanted feature of negative 'expected spike sizes' in most cases. However, still some of the low prices are classified as being in the spike regime, see the lower left panel in Figure 1. The lower number of 'sudden drops' classified as spikes suggests that the calibration scheme does a better job of identifying the spikes in prices than in log-prices. But the classification is far from perfect.

5.2. 2-regime models with shifted spike distributions

To cope with the problem of spike misclassification, Janczura and Weron (2009) introduced median-shifted lognormal:

$$\log(X_{t,s} - X(0.5)) \sim \mathrm{N}(\mu_s, \sigma_s^2), \quad X_{t,s} > X(0.5), \tag{10}$$

and Pareto:

$$X_{t,s} \sim F_{\mathrm{Pareto}}(\sigma_s, \mu_s) = 1 - \left(\frac{\mu_s}{x}\right)^{\sigma_s}, \quad x > \mu_s \geq X(0.5), \tag{11}$$

spike regime distributions. Here $X(0.5)$ denotes the median of the dataset (deseasonalized prices or log-prices), but more generally it can be any quantile $X(q)$, $q \in (0, 1)$. Note, that shifted spike distributions assign zero probability to prices below a certain quantile of the dataset.

The goodness-of-fit statistics for the 2-regime models with shifted spike distributions fitted to the six considered datasets are summarized in Table 1. They include the quantile-based measures IQR and IDR, the log-likelihoods and the p-values of the K-S test for the individual regimes and the whole model (for computational details see Appendix B). The reported IQR and IDR values are in fact average (over 100 samples) percentage changes between the values of IQR and IDR for the dataset and the simulated trajectories. A positive value, say, 9% for EEX1 (first row in Table 1) indicates that the model implied IQR is 9% wider than the IQR of the deseasonalized data. A negative value, say, −3% for PJM2 (fourth row in Table 1) means that the IQR of the deseasonalized data is 3% wider than the model implied IQR.

The IQR and IDR measures imply that while the models more or less capture the quantiles of the price processes, they fail in case of log-prices. This is especially true for the most spiky EEX market (in both periods). Only one of the models for log-prices – median-shifted lognormal spikes for PJM2 – passes the K-S test at the 5% level. But this is the dataset with the lowest number of extreme observations, see the middle right panel in Figure 7.

Comparing the lognormal and Pareto spikes we see that the latter do not provide a good fit to the data. None of the models with Pareto spikes passes the K-S test for the spike regime (except for PJM2 log-prices), although in one case (PJM1 prices) the overall model fit is acceptable. On the other hand, the lognormal spike regime itself passes the K-S test in all cases (both for prices and log-prices). As a result, the models with median-shifted lognormal spikes yield a reasonable fit to moderately spiky prices in the PJM and NEPOOL markets. Comparing Figure 4 with the bottom panels in Figure 2 we can also observe that the fits of the median-shifted lognormal spike regime distribution functions to the empirical ones are much better than for the models with non-shifted lognormal spikes.

Looking at the log-likelihoods we can observe a similar picture. In all cases (except for EEX1 log-prices) the models with shifted lognormal spikes provide a better fit than the corresponding models with shifted Pareto spikes.

Table 1: Goodness-of-fit statistics for 2-regime models with Vasicek, see eqns. (5)-(6), base regime dynamics and median-shifted lognormal or Pareto spike distributions. Models for prices are summarized in columns 2-7, for log-prices in columns 8-13. p-values of 0.05 or more are emphasized in bold.

	Prices						Log-prices					
	Simulation			K-S test p-value			Simulation			K-S test p-value		
Data	IQR	IDR	LogL	Base	Spike	Model	IQR	IDR	LogL	Base	Spike	Model
					Shifted lognormal spikes							
EEX1	9%	11%	-4193.7	0.0012	**0.4061**	0.0032	30%	30%	403.0	0.0000	**0.7371**	0.0000
EEX2	13%	3%	-5066.9	0.0090	**0.4732**	0.0149	27%	16%	399.6	0.0000	**0.6313**	0.0000
PJM1	9%	-1%	-4385.9	0.0341	**0.4346**	**0.0530**	19%	8%	777.4	0.0007	**0.3219**	0.0012
PJM2	-3%	3%	-5012.1	**0.0887**	**0.9196**	**0.0893**	3%	6%	780.1	**0.0747**	**0.4147**	**0.0696**
NEP1	2%	2%	-4327.1	0.0247	**0.5093**	**0.0561**	9%	8%	610.4	0.0002	**0.8316**	0.0003
NEP2	0%	-2%	-4665.9	**0.0823**	**0.8416**	**0.1251**	8%	0%	1417.9	0.0088	**0.7430**	0.0170
					Shifted Pareto spikes							
EEX1	7%	9%	-4218.6	0.0000	0.0000	0.0000	27%	27%	436.9	0.0000	0.0123	0.0000
EEX2	10%	1%	-5101.8	0.0188	0.0008	0.0412	26%	16%	374.6	0.0000	0.0166	0.0000
PJM1	8%	-5%	-4447.2	**0.0500**	0.0000	**0.0500**	20%	6%	755.2	0.0007	0.0000	0.0012
PJM2	0%	1%	-5161.6	0.0041	0.0000	0.0007	6%	7%	744.9	0.0262	**0.3508**	0.0147
NEP1	-2	-6%	-4366.1	0.0300	0.0000	0.0300	8%	3%	546.7	0.0001	0.0000	0.0003
NEP2	13%	0%	-4703.1	0.0230	0.0000	0.0097	9%	0%	1409.7	0.0024	0.0000	0.0083

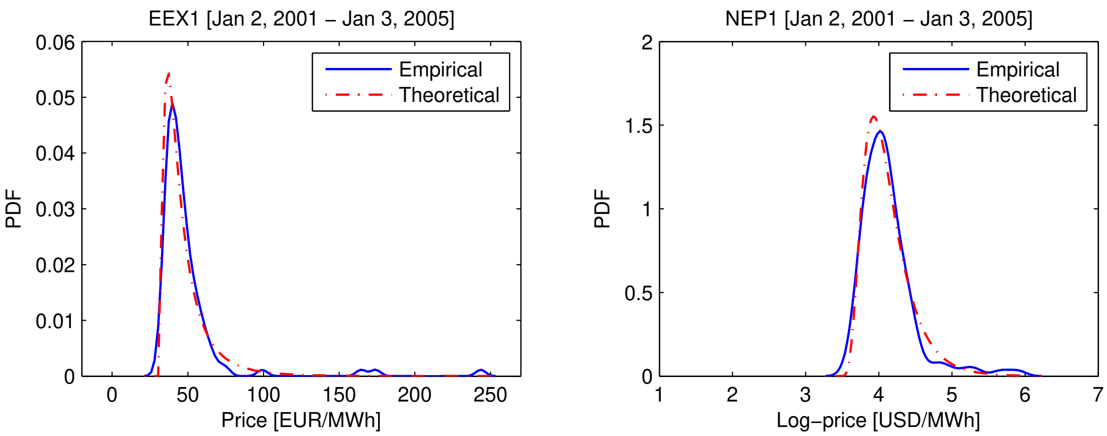

Figure 4: Comparison of empirical (sample) and theoretical (model implied) spike regime probability distribution functions in the 2-regime model with median-shifted lognormal spikes and Vasicek base regime dynamics. The fits are much better than for the models with non-shifted spike regime distributions, see Figure 2.

This is somewhat surprising given that distributions with power-law decay (α-stable) provide a better fit to the deseasonalized EEX and NEPOOL price changes than the lighter tailed distributions (hyperbolic, NIG), see Weron (2009). Perhaps, the better fit of a power-law distribution to the few extreme observations does not offset the worse fit to the less severe prices in the considered models.

5.3. 2-regime models with shifted spike distributions and heteroscedastic base regime dynamics

Due to the cutoff at the median in the models with median-shifted spike regime distributions, the price or log-price 'drops' are not classified as spikes anymore. This does not, however, seem to be a good approach for log-price models (except for datasets with practically no 'drops' as PJM2); the extremely low log-prices can hardly be modeled by the Vasicek process. A possible remedy is to use different dynamics for the base regime. Janczura and Weron (2009) utilized the square root process of Cox et al. (1985), but nothing prevents us from considering here more general heteroscedastic processes of the form:

$$dX_{t,b} = (\alpha - \beta X_{t,b})dt + \sigma_b X_{t,b}^{\gamma} dW_t, \tag{12}$$

Table 2: Goodness-of-fit statistics for 2-regime models with heteroscedastic base regime dynamics and median-shifted lognormal spike distributions. Models for prices are summarized in columns 2-8, for log-prices in columns 9-15. p-values of 0.05 or more are emphasized in bold.

	Prices							Log-prices						
		Simulation			K-S test p-value				Simulation			K-S test p-value		
Data	γ	IQR	IDR	LogL	Base	Spike	Model	γ	IQR	IDR	LogL	Base	Spike	Model
							Shifted lognormal spikes							
EEX1	-0.43	0%	0%	-4169.3	0.0022	**0.2365**	0.0050	-4.08	22%	26%	625.5	0.0000	**0.9865**	0.0000
EEX2	-0.32	10%	2%	-5041.7	0.0125	**0.2306**	0.0276	-3.69	22%	12%	551.8	0.0000	**0.5875**	0.0000
PJM1	0.10	5%	1%	-4356.4	**0.0853**	**0.5408**	**0.1607**	-1.02	17%	6%	793.1	0.0006	**0.1924**	0.0011
PJM2	0.16	1%	-1%	-4989.3	**0.5882**	**0.1802**	**0.5435**	-0.01	1%	2%	804.2	**0.0582**	**0.1843**	**0.0995**
NEP1	0.22	2%	0%	-4326.3	0.0317	**0.4754**	**0.0742**	-1.35	9%	12%	643.1	0.0003	**0.8524**	0.0003
NEP2	0.62	0%	0%	-4654.0	**0.0828**	**0.3566**	**0.0983**	-2.37	1%	-1%	1445.2	0.0368	**0.1724**	**0.0980**

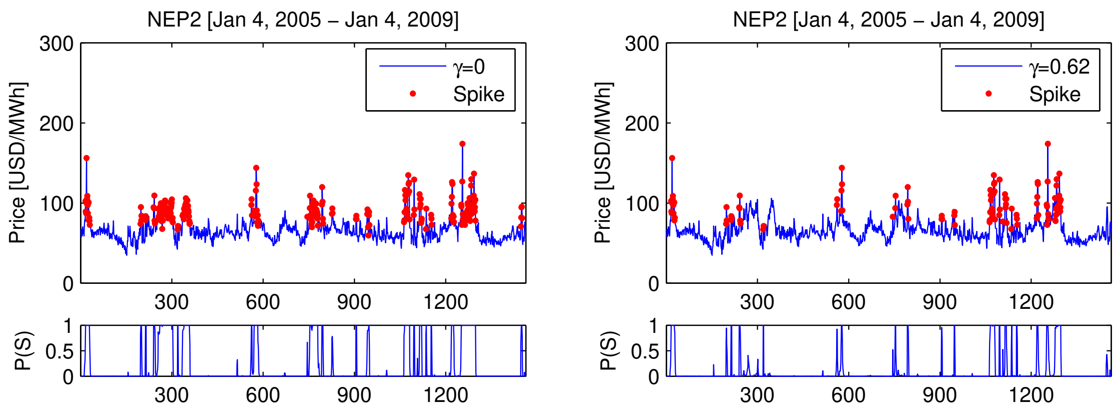

Figure 5: Sample calibration results for the 2-regime model with median-shifted lognormal spikes fitted to NEP2 prices. The difference between Vasicek (*left*) and heteroscedastic (*right*) base regime dynamics is clearly visible. Note, that due to the cutoff at the median, none of the price 'drops' are classified as spikes anymore.

where γ = const. Note, that in this model the volatility is dependent on the current price level $X_{t,b}$. For positive γ, the higher the price level the larger are the price changes. For negative γ, the lower the price level the larger are the price changes. Compared to Vasicek dynamics (5)-(6), we can expect that in this model the moderately extreme prices will be classified as 'normal' and not spiky. Indeed, this effect can be observed in Figure 5. The corresponding discrete time process:

$$X_{t,b} = \alpha - (1-\beta)X_{t-1,b} + \sigma_b X_{t-1,b}^{\gamma} \epsilon_t, \tag{13}$$

is obtained by applying the Euler scheme to (12).

Calibration results for the 2-regime models with median-shifted lognormal spikes and heteroscedastic base regime dynamics are presented in Figures 6 and 7. The goodness-of-fit statistics are summarized in Table 2. Results for the models with Pareto spikes are not reported due to the poor fit, even after allowing for $\gamma <> 0$. Only lognormal spike distributions will be considered in the remainder of the paper.

Comparing the log-likelihoods with the ones for Vasicek base regime models, we note that in all cases a better fit was obtained. The p-values of the K-S goodness-of-fit test have not, however, increased in all cases. This is likely due to the fact that the K-S test focuses on the largest deviation between the model and empirical distribution function, while the log-likelihood averages over all observations. Probably using the integral-type W^2 statistic of Cramér-von Mises (see e.g. Čižek et al., 2005) instead of the extremum-based Kolmogorov-Smirnov statistic would result in more convergent behavior of the two goodness-of-fit measures.

Looking at the IQR and IDR measures we see a similar picture to the one in Table 1. Although the deviations from dataset statistics are generally smaller than before, in most cases the models still fail to capture the quantiles of the log-prices. Again this feature is most pronounced for the spiky EEX log-prices. Also EEX prices are the hardest to model. While the considered models yield a reasonable fit to moderately spiky prices in the PJM and NEPOOL markets, the EEX base regime dynamics does not conform to the model implied price dynamics. Note, that in the case of this market the estimated γ is negative. This suggests that the base regime model (12) tries to catch the lower than

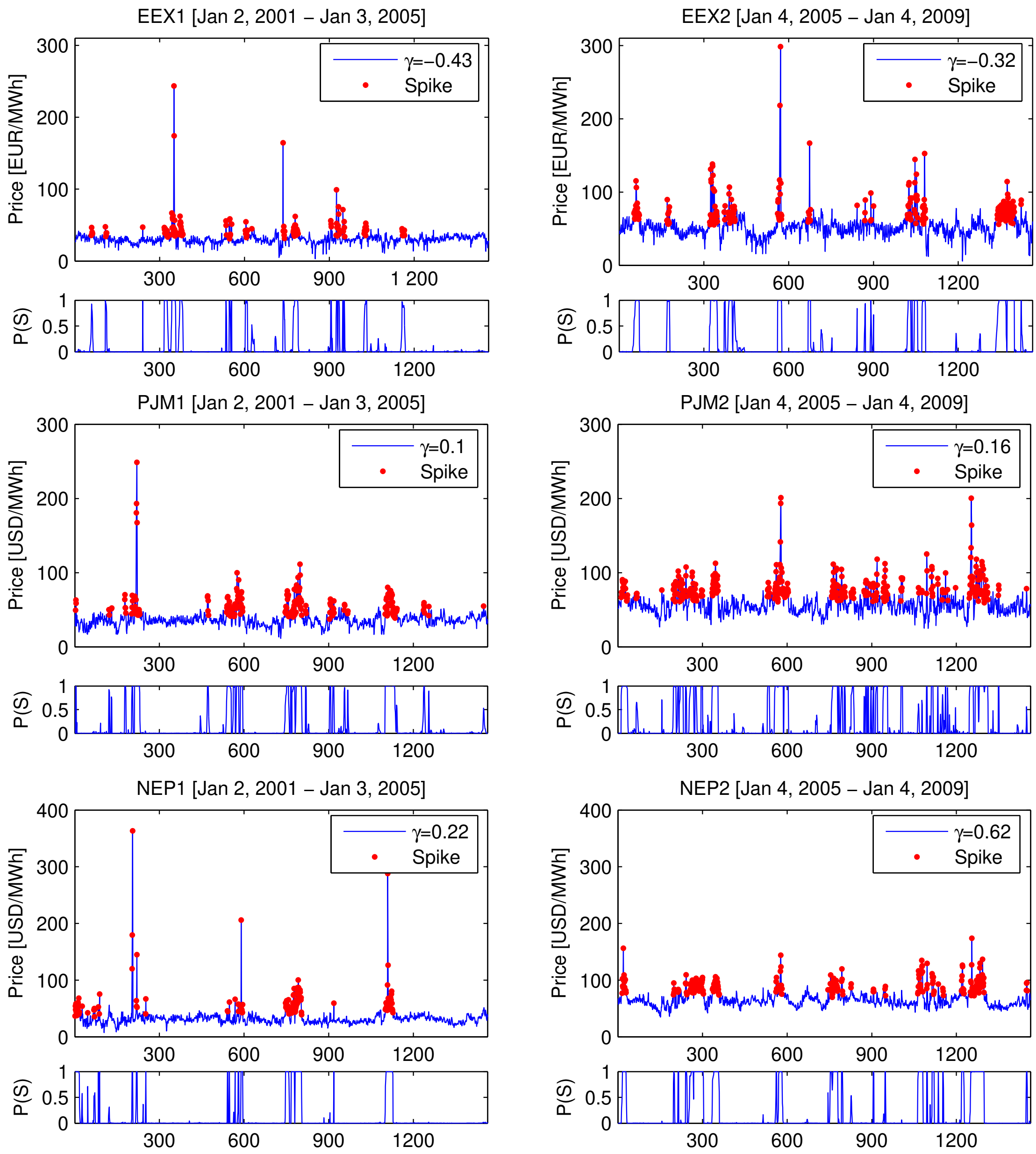

Figure 6: Calibration results for the 2-regime model with median-shifted lognormal spikes and heteroscedastic base regime dynamics fitted to prices.

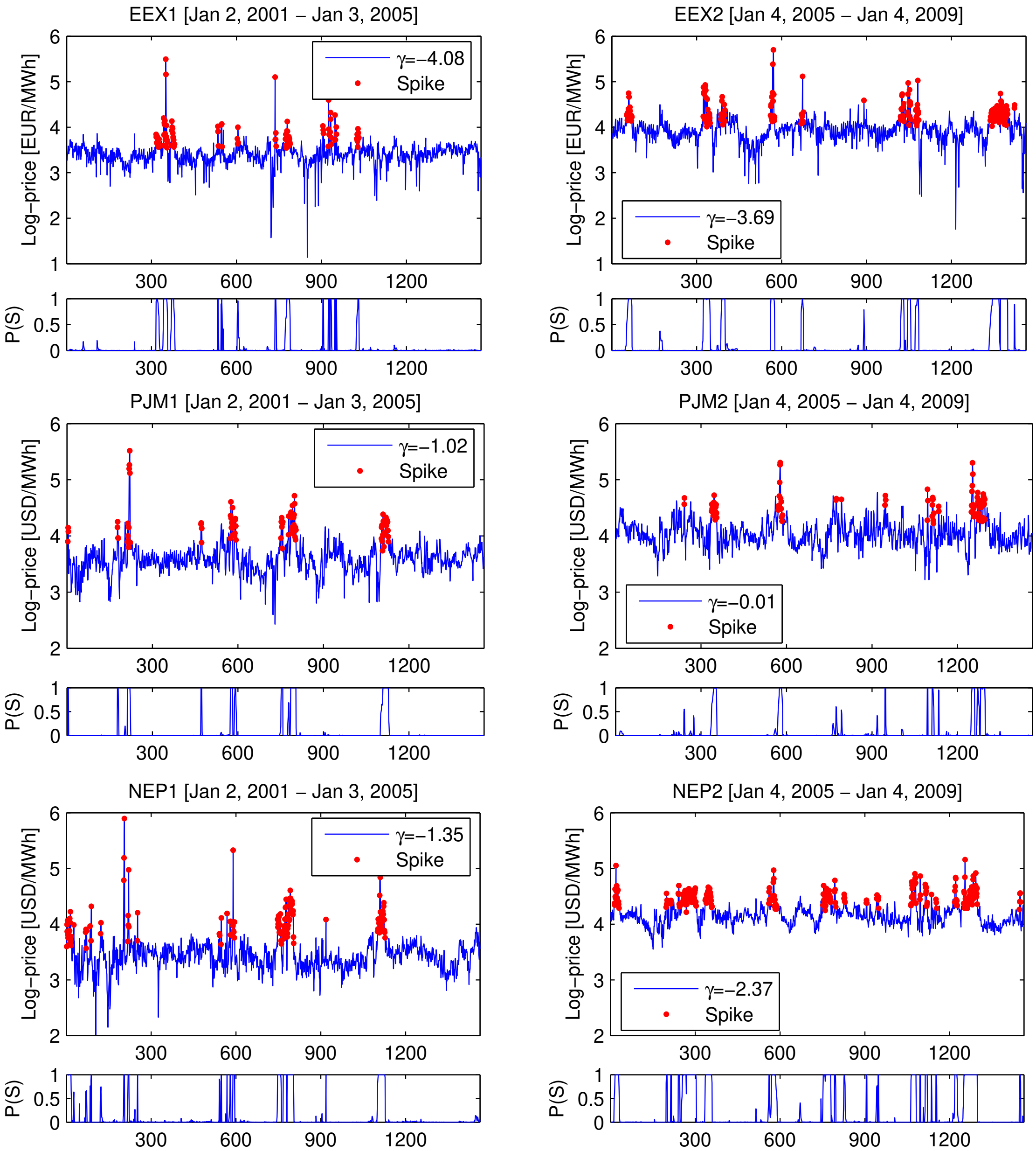

Figure 7: Calibration results for the 2-regime model with median-shifted lognormal spikes and heteroscedastic base regime dynamics fitted to log-prices.

Table 3: Calibration results for the IS 3-regime model with heteroscedastic base regime dynamics and median-shifted lognormal spikes and drops. Parameter estimates are summarized in columns 2-9, transition probabilities for staying in the same regime p_{ii} in columns 10-12 and unconditional probabilities of being in each of the regimes in columns 13-15.

	Parameters								Probabilities					
Data	γ	α	β	σ_b^2	μ_s	σ_s^2	μ_d	σ_d^2	p_{bb}	p_{ss}	p_{dd}	$R_t = b$	$R_t = s$	$R_t = d$
							Prices							
EEX1	0.6309	14.0717	0.4509	0.1215	2.4027	0.5590	1.7227	0.3573	0.91	0.80	0.81	0.67	0.11	0.22
EEX2	0.3070	17.8797	0.3501	3.2502	2.9152	0.5562	2.8504	0.1800	0.96	0.90	0.86	0.74	0.17	0.09
PJM1	0.6595	9.4038	0.2607	0.1232	2.9057	0.4640	2.4766	0.0967	0.95	0.82	0.79	0.78	0.13	0.09
PJM2	0.1724	21.0813	0.3792	7.9227	3.0481	0.3134	2.6768	0.0949	0.88	0.83	0.73	0.63	0.23	0.14
NEP1	0.5262	7.9267	0.2587	0.4608	3.2267	0.4883	2.7220	0.0689	0.98	0.84	0.73	0.91	0.08	0.01
NEP2	0.0742	13.0526	0.2063	13.8627	3.1856	0.2459	2.5783	0.1033	0.97	0.91	0.92	0.76	0.18	0.06
							Log-prices							
EEX1	0.4102	1.5442	0.4491	0.0034	-1.1627	0.3322	-1.5900	0.5313	0.91	0.80	0.81	0.66	0.10	0.24
EEX2	1.9558	0.8815	0.2445	0.0001	-0.5775	0.2220	-0.9674	0.2025	0.98	0.90	0.84	0.85	0.06	0.10
PJM1	0.4481	0.9874	0.2877	0.0054	-0.4987	0.2135	-0.9753	0.1921	0.97	0.86	0.79	0.84	0.08	0.09
PJM2	0.5057	1.0746	0.2663	0.0037	-0.8978	0.1659	-1.0892	0.1483	0.96	0.84	0.78	0.80	0.11	0.08
NEP1	1.9984	0.8651	0.2128	0.0001	-0.7526	0.1151	-0.9654	0.1328	0.99	0.63	0.82	0.94	0.01	0.05
NEP2	1.0923	0.9094	0.2184	0.0003	-1.1018	0.1540	-1.5246	0.1589	0.96	0.91	0.91	0.70	0.16	0.14

Table 4: Goodness of fit statistics for the IS 3-regime models with heteroscedastic base regime dynamics and median-shifted lognormal spikes and drops. p-values of 0.05 or more are emphasized in bold.

	Simulation			K-S test p-values			
Data	IQR	IDR	LogL	Base	Spike	Drop	Model
				Prices			
EEX1	-1%	3%	-3798.2	**0.8371**	**0.0726**	**0.8576**	**0.5719**
EEX2	5%	-1%	-4848.5	**0.5510**	**0.2920**	**0.9196**	**0.3168**
PJM1	2%	0%	-4153.9	**0.3876**	**0.7052**	**0.7715**	**0.4072**
PJM2	2%	0%	-4723.2	**0.4824**	**0.3273**	0.0244	**0.4828**
NEP1	0%	0%	-4266.6	0.0404	**0.6121**	**0.9092**	**0.0609**
NEP2	5%	0%	-4610.3	**0.1359**	**0.7911**	**0.8771**	**0.1059**
				Log-prices			
EEX1	-2%	5%	1181.2	**0.7297**	**0.3341**	**0.1971**	**0.5001**
EEX2	7%	1%	931.1	**0.1392**	**0.3571**	**0.2095**	**0.4234**
PJM1	8%	1%	1002.9	**0.3413**	**0.2726**	**0.9080**	**0.4640**
PJM2	1%	2%	1030.2	**0.2165**	**0.4604**	**0.2887**	**0.5258**
NEP1	1%	2%	864.1	**0.1661**	**0.8762**	**0.2947**	**0.2553**
NEP2	6%	0%	1573.7	**0.6858**	**0.7338**	**0.2334**	**0.8796**

'normal' prices observed in the data, leaving the extreme positive observations to be modeled by the spike regime. For the same reason γ is negative in case of log-price models. This behavior is the main motivation for considering 3-regime models in the next section.

5.4. A new class of 3-regime models

The 3-regime model of Huisman and Mahieu (2003) assumed that the initial jump regime was immediately followed by the reversing regime and then moved back to the base regime, i.e. $p_{sb} = p_{db} = 1$ in the transition matrix (2). In doing so, their model did not allow for consecutive high prices (in fact of log-prices). We see no reason for this limitation and in this section consider independent spike (IS) 3-regime models with heteroscedastic base regime dynamics of the form (12), shifted lognormal distribution (10) for the spike regime and inverted shifted lognormal distribution:

$$\log(-X_{t,d} + X(0.5)) \sim \mathrm{N}(\mu_d, \sigma_d^2), \quad X_{t,d} < X(0.5), \tag{14}$$

for the 'drop' (or 'downward spike') regime. The 'inversion' can be interpreted as taking a mirror image of the lognormal probability density function with respect to the origin.

Calibration results for the IS 3-regime models with heteroscedastic base regime dynamics and median-shifted lognormal spikes and drops are presented in Figures 8 and 9. Parameter estimates and goodness-of-fit statistics are summarized in Tables 3 and 4, respectively.

Comparing with Table 2 we note that the parameter γ is no longer negative. The 'drop' regime seems to do the job of modeling the low (log-)prices well. For price models γ ranges from 0.0742 (nearly Vasicek dynamics) for NEP2 to 0.6595 (a little over CIR-type dynamics) for PJM1. For log-price models γ is generally much higher – from 0.4102 (nearly CIR-type dynamics) for EEX1 to 1.9984 for NEP1. The high γ results in the base regime covering most of the moderately spiky log-prices – the probability of being in the spike regime $P(R_t = s)$ is only 1% for NEP1 log-prices. This feature is also visible in the bottom left panel of Figure 9. The unconditional probabilities for price and log-price models are qualitatively similar. The probability of being in the base regime $P(R_t = b)$ ranges from 0.63 to 0.91 for prices and from 0.66 to 0.94 for log-prices. The probabilities of being in one of the extreme regimes are generally significantly lower, however, EEX1 drops and PJM2 spikes are likely to be observed for more than 20% of time. The probabilities of staying in a given regime p_{ii} are relatively high, with p_{bb} being the highest and ranging from 0.88 for PJM2 prices to 0.99 for NEP1 log-prices. The spikes tend to be more persistent than drops – the probabilities p_{ss} are on average higher than p_{dd} – but the differences are not large.

Regarding the robustness of the results in the two considered periods (2001-2004 vs. 2005-2008) we note that, while in the case of log-price models there is no recognizable pattern, the price models in the second period are less heteroscedastic. They exhibit lower values of γ which is offset by higher volatility σ_b^2 in the base regime. Also the level parameter α is higher in the second period. But this is due to the deseasonalization scheme which shifts the prices so that the minimum of the new process is the same as the minimum of the original prices.

Comparing the log-likelihoods with the ones for 2-regime models, we note that in all cases a better fit was obtained. In fact, in most cases a significantly better fit. However, not all the p-values of the K-S goodness-of-fit test have increased. This is likely due to the fact that the K-S test focuses on the largest deviation between the model and empirical distribution function, while the (log)likelihood averages over all observations. Probably using an integral-type statistic, like the W^2 of Cramér-von Mises (see e.g. Čižek et al., 2005), instead of the extremum-based Kolmogorov-Smirnov statistic would result in more convergent behavior of the two goodness-of-fit measures.

Also looking at the IQR and IDR measures we see a general improvement over the results for previously analyzed models (Tables 1-2). The deviations for price models do not exceed 5%. For log-price models the improvement is even more visible – the deviations do not exceed 8%, whereas before they reached as much as 27% in Table 2 and 30% in Table 1.

The EEX (log-)prices are no longer hard to model. All corresponding p-values are well above the 5% threshold. Now, the PJM2 prices exhibit the lowest fit ... but only for the drop regime (p-value of 0.0244); the whole model p-value is nearly 50%. Overall all models yield acceptable fits. The situation is even better for log-price models. Here not only model p-values but also all p-values for the individual regimes exceed the 5% threshold.

6. Conclusions

In this paper we have calibrated and tested a range of Markov regime switching (MRS) models in an attempt to find parsimonious specifications that not only address the main characteristics of electricity prices but are statistically sound as well. To this end, we have applied not only the standard descriptive statistics, but also performed hypothesis testing (based on the Kolmogorov-Smirnov test). This approach allowed us to identify correctly specified MRS models.

The analysis of models proposed in the literature has revealed their weakness. While most of the models are elegant, their fit to empirical data is often statistically unacceptable. This situation has led us to proposing in Section 5.4 a new class of independent spike (IS) 3-regime models with shifted spike distributions and heteroscedastic base regime dynamics. In contrast to the 3-regime model of Huisman and Mahieu (2003) we allow for consecutive spikes (high prices) or drops (low prices), which makes the model more universal.

This new class of models provides a statistically acceptable fit to all six datasets. Also visually the fit acceptable. However, in a few cases there seem to be too many (log-)prices classified into one of the extreme price regimes. In particular, EEX1 drops and PJM2 spikes are likely to be observed for more than 20% of time. We believe that there is still room for improvement in this case. A possible remedy would be to shift the spike and drop distributions not by the median, but by a different (higher) quantile of the dataset. Indeed, preliminary estimation and simulation results seem to confirm this hypothesis, however, more research is needed to find a robust and universal solution.

Appendix A. Estimation procedure

The estimation procedure, which draws on Kim (1994), starts with an arbitrarily chosen vector of initial parameters $\theta^{(0)} = (\alpha^{(0)}, \beta^{(0)}, \sigma_b^{(0)}, \gamma^{(0)}, \mu_s^{(0)}, \sigma_s^{(0)}, \mu_d^{(0)}, \sigma_d^{(0)}, \mathbf{P}^{(0)})$. For parameter definitions see eqns. (7)-(14). Note, that for the

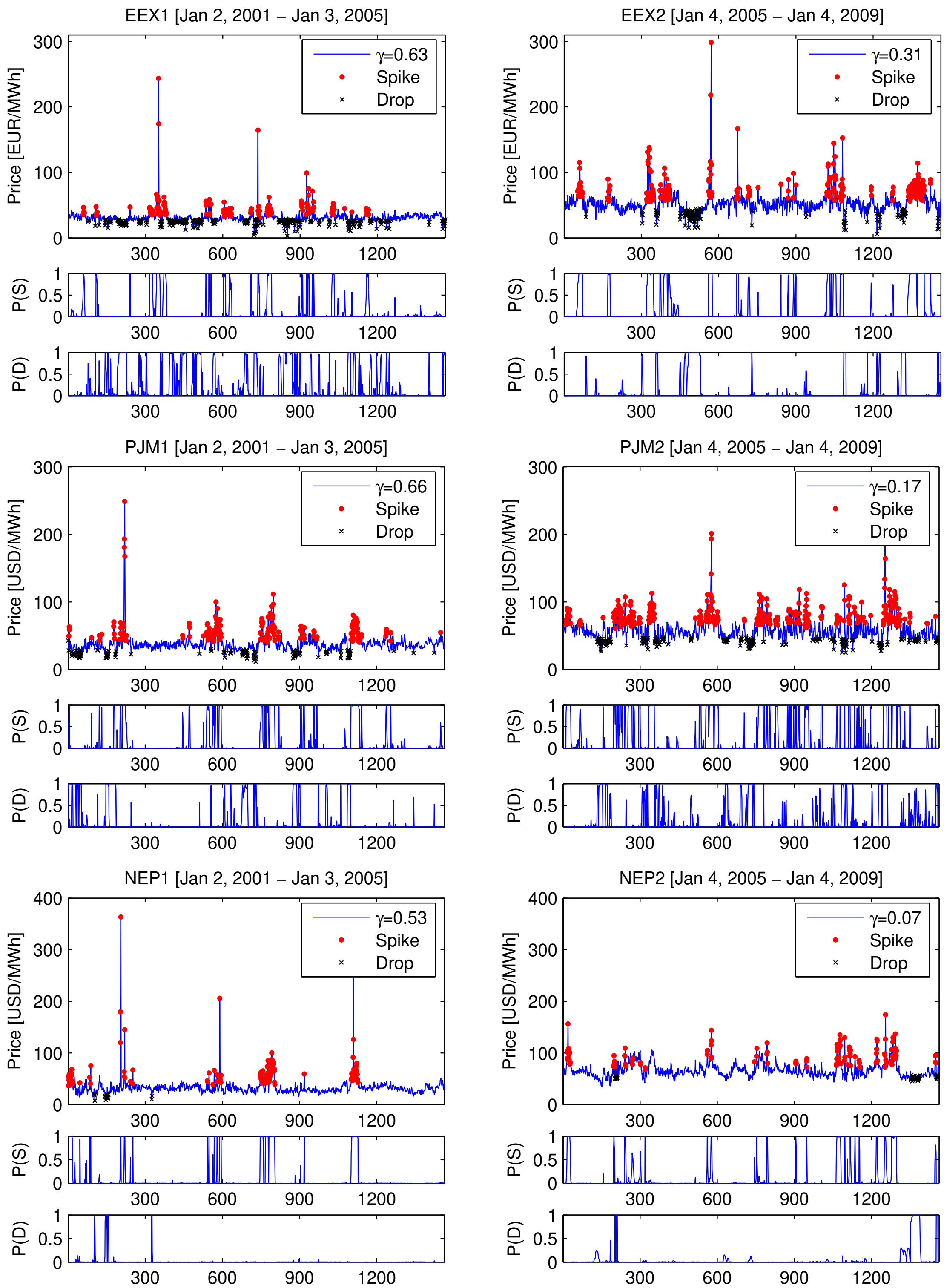

Figure 8: Calibration results for the IS 3-regime models with heteroscedastic base regime dynamics and median-shifted lognormal spikes and drops fitted to prices.

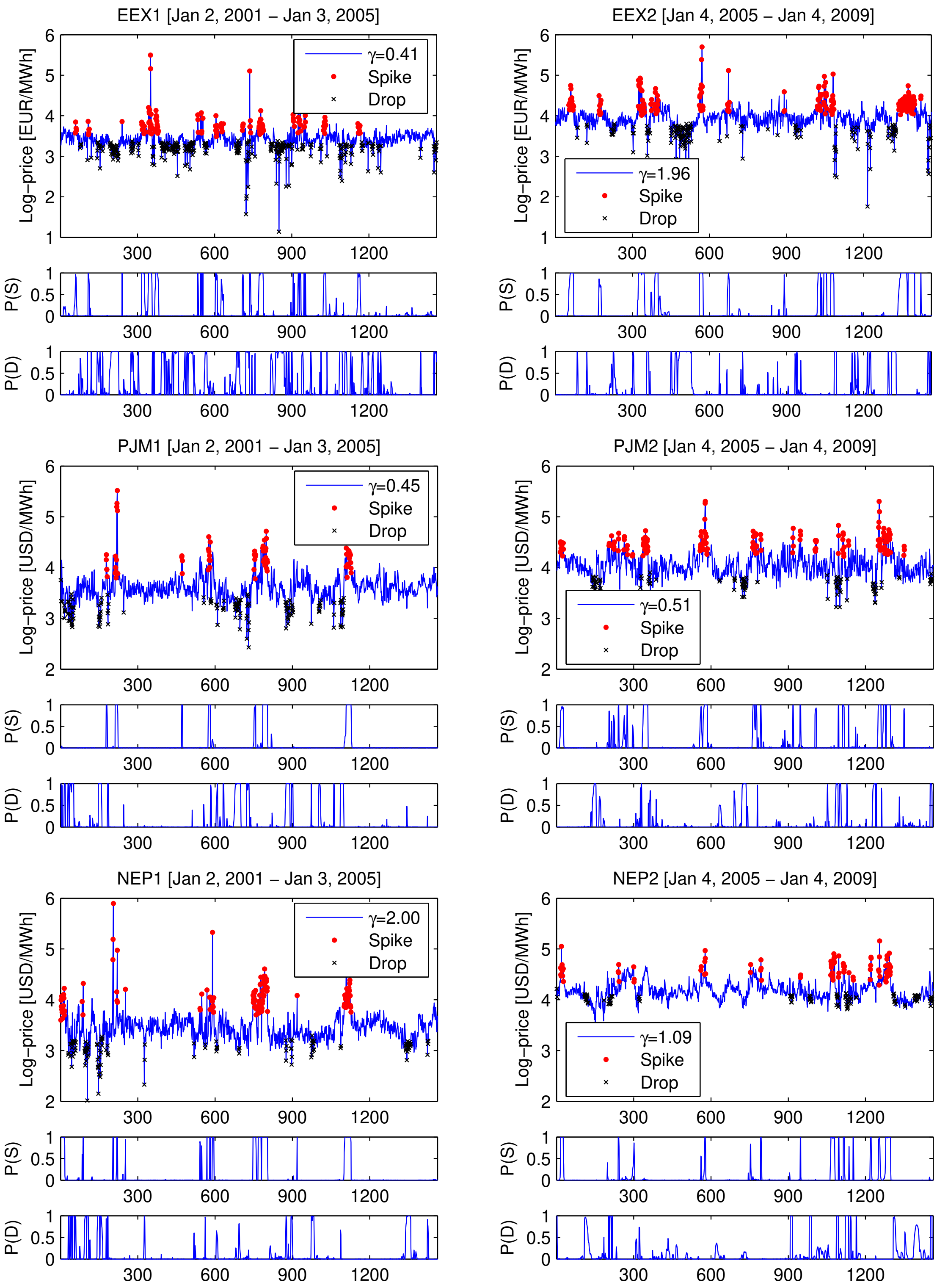

Figure 9: Calibration results for the IS 3-regime models with heteroscedastic base regime dynamics and median-shifted lognormal spikes and drops fitted to log-prices.

2-regime models the 'drop' regime does not exist and, hence, parameters μ_d and σ_d are not defined. These values are then used to calculate the conditional probabilities $P(R_t = i|x_1, ..., x_T; \theta^{(0)})$ for the process X_t being in regime $R_t = i$ at time t, the so-called 'smoothed inferences'. In the case when the value of the process depends at most on the last observation, X_{t-1}, the algorithm consists of the following steps:

i) for $t = 1, 2, ..., T$ iterate on equations:

$$P(R_t = i|x_1, ..., x_t) = \frac{P(R_t = i|x_1, ..., x_{t-1}) f(x_t|R_t = i; x_1, ..., x_{t-1})}{\sum_{i\in\{b,s,d\}} P(R_t = i|x_1, ..., x_{t-1}) f(x_t|R_t = i; x_1, ..., x_{t-1})}, \tag{A.1}$$

and

$$P(R_{t+1} = j|x_1, ..., x_t) = \sum_{j\in\{b,s,d\}} p_{ij}^{(0)} P(R_t = i|x_1, ..., x_t), \tag{A.2}$$

where $f(x_t|R_t = i; x_1, ..., x_{t-1})$ is the density of the underlying process at time t, conditional on being in regime $i \in \{b, s, d\}$. Note, that the parameter vector, θ, is omitted in $P(\cdot|\cdot; \theta)$ to simplify the notation.

ii) for $t = T, T-1, ..., 1$ iterate on

$$P(R_t = i|x_1, ..., x_T) = \sum_{i\in\{b,s,d\}} \frac{P(R_t = i|x_1, ..., x_t) P(R_{t+1} = j|x_1, ..., x_T) p_{ij}^{(0)}}{P(R_{t+1} = i|x_1, ..., x_t)}. \tag{A.3}$$

The crucial point in the above procedure is to derive $f(x_t|R_t = i; x_1, ..., x_{t-1})$ for the regime in which the value of the process X_t depends on the last observation x_{t-1}. In the models considered here, this is true for the base regime only. If x_t, as well as, x_{t-1} come from the base regime, then as a result of definition (13), X_t has a conditional Gaussian distribution given by the formula:

$$f(x_t|R_t = b, R_{t-1} = b; x_1, ..., x_{t-1}) = \frac{1}{\sqrt{2\pi}\sigma_b^{(0)} x_{t-1}^{\gamma^{(0)}}} \exp\left(-\frac{(x_t - (1-\beta^{(0)})x_{t-1} - \alpha^{(0)})^2}{2(\sigma_b^{(0)})^2 x_{t-1}^{2\gamma^{(0)}}}\right). \tag{A.4}$$

However, since observations from different regimes are independent and during a spike or a drop the base regime becomes latent, formula (A.4) cannot be directly applied. Instead, $f(x_t|R_t = b, R_{t-1} \neq b, ..., R_{t-k} \neq b, R_{t-k-1} = b; x_1, ..., x_{t-1})$ and the whole set of probabilities $P(R_t = i_t, R_{t-1} = i_{t-1}, ..., R_{t-k} = i_{t-k})$ should be used in steps i)-ii). Obviously, this leads to a high computational burden. As a solution, Huisman and de Jong (2003) suggested to use probabilities of only the last 10 observations, while Ethier and Mount (1998) used a 2-state model with mean-reverting AR(1) processes sharing the same set of random innovations (only the parameters of the two AR(1) processes varied between the regimes).

Instead, we propose to replace the latent variables from the base regime with their expectations $E(\hat{X}_{t,b}|x_t, x_{t-1}, ..., x_{t-k})$, where x_{t-k} is the last observation coming from the base regime with probability equal one (i.e. $P(R_{t-k} = b) = 1$ and $P(R_j = b) < 1$ for $t > j > t-k$). The subscripts in $X_{t,b}$ denote the time and regime, respectively. These expectations can be obtained from the following recursion:

1. for $k = 1$ calculate

$$E(\hat{X}_{t,b}|x_t, x_{t-1}) = x_t P(R_t = b|x_t) + (\alpha^{(0)} + (1-\beta^{(0)})x_{t-1}) P(R_t \neq b|x_t)$$

2. for $k > 1$ calculate

$$E(\hat{X}_{t,b}|x_t, x_{t-1}, ..., x_{t-j}) = x_t P(R_t = b|x_t) + (\alpha^{(0)} + (1-\beta^{(0)}) E(\hat{X}_{t-1,b}|x_{t-1}, ..., x_{t-k})) P(R_t \neq b|x_t).$$

Now, formula (A.4) with $E(\hat{X}_{t,b}|x_t, x_{t-1}, ..., x_{t-k})$ instead of x_t can be used and steps i)-ii) require storing probabilities concerning the last observed price only.

In the second step of the EM algorithm, new and more exact maximum likelihood (ML) estimates $\hat{\theta}^{(1)}$ for all model parameters are calculated. Compared to standard ML estimation, where for a given probability density function f the log-likelihood function $\sum_{t=1}^{T} \log f(x_t, \theta)$ is maximized, here each component of this sum has to be weighted with the corresponding smoothed inference, since each observation x_t belongs to the ith regime exactly with probability $P(R_t = i|x_1, ..., x_T)$.

From the discrete time representation (13) of the mean-reverting model for the base regime we have that the conditional density of x_t given x_{t-1} is Gaussian with mean $\alpha - (1-\beta)x_{t-1}$ and standard deviation σx_{t-1}^{γ}. Since the joint density $f(x_1, x_2, ..., x_T)$ can be written as a product of appropriate conditional densities

$$f(x_T|x_{T-1}, x_{T-2}, ..., x_1) f(x_{T-1}|x_{T-2}, x_{T-3}, ..., x_1) \ldots f(x_2|x_1) f(x_1),$$

the log-likelihood function is given by the following formula:

$$\ln[L(\alpha, \beta, \sigma_b)] = \sum_{t=2}^{T} P(R_t = b) \left[\ln\left(\frac{1}{\sqrt{2\pi}\sigma_b x_{t-1}^{\gamma}} \right) - \frac{(x_t - (1-\beta)x_{t-1} - \alpha)^2}{2\sigma_b^2 x_{t-1}^{2\gamma}} \right]. \quad \text{(A.5)}$$

In order to find the ML estimates the partial derivatives of $\ln(L)$ are set to zero. This leads to the following formulas for the estimates of the base regime:

$$\hat{\beta} = \frac{\sum_{t=2}^{T} P(R_t = b) x_{t-1}^{1-2\gamma} \left(x_t - x_{t-1} - \frac{\sum_{t=2}^{T} P(R_t=b) x_{t-1}^{-2\gamma}(x_t - x_{t-1})}{\sum_{t=2}^{T} P(R_t=b) x_{t-1}^{-2\gamma}} \right)}{\sum_{t=2}^{T} P(R_t = b) x_{t-1}^{1-2\gamma} \left(\frac{\sum_{t=2}^{T} P(R_t=b) x_{t-1}^{1-2\gamma}}{\sum_{t=2}^{T} P(R_t=b) x_{t-1}^{-2\gamma}} - x_{t-1} \right)},$$

$$\hat{\alpha} = \frac{\sum_{t=2}^{T} P(R_t = b) x_{t-1}^{-2\gamma}(x_t - (1-\hat{\beta})x_{t-1})}{\sum_{t=2}^{T} P(R_t = b) x_{t-1}^{-2\gamma}},$$

$$\hat{\sigma_b^2} = \frac{\sum_{t=2}^{T} P(R_t = b) x_{t-1}^{-2\gamma}(x_t - \hat{\alpha} - (1-\hat{\beta})x_{t-1})^2}{\sum_{t=2}^{T} P(R_t = b)}.$$

Parameter γ requires numerical maximization of the likelihood function. In the above formulas, latent variables from the base regime are replaced by their expectations.

The maximum likelihood estimates of the spike regime parameters are equal to:

$$\hat{\mu_s} = \frac{\sum_{t=1}^{T} \ln(X_t - X(q)) \mathbb{I}_{\{X_t > X(q)\}} P(R_t = s)}{\sum_{t=1}^{T} P(R_t = s)}, \quad \text{(A.6)}$$

$$\hat{\sigma_s^2} = \frac{\sum_{t=1}^{T} (\ln(X_t - X(q)) - \hat{\mu_s})^2 \, \mathbb{I}_{\{X_t > X(q)\}} P(R_t = s)}{\sum_{t=1}^{T} P(R_t = s)}. \quad \text{(A.7)}$$

For the 3-regime models, the drop regime parameters are given by:

$$\hat{\mu_d} = \frac{\sum_{t=1}^{T} \ln(-X_t + X(q)) \mathbb{I}_{\{X_t < X(q)\}} P(R_t = d)}{\sum_{t=1}^{T} P(R_t = d)}, \quad \text{(A.8)}$$

$$\hat{\sigma_d^2} = \frac{\sum_{t=1}^{T} (\ln(-X_t + X(q)) - \hat{\mu_d})^2 \, \mathbb{I}_{\{X_t < X(q)\}} P(R_t = d)}{\sum_{t=1}^{T} P(R_t = d)}. \quad \text{(A.9)}$$

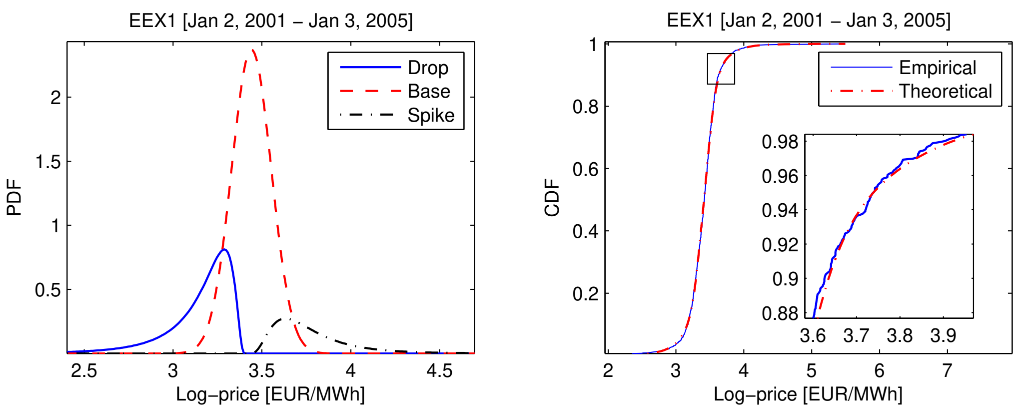

Figure B.10: Sample calibration details for an IS 3-regime model with median-shifted lognormal distribution and heteroscedastic base regime dynamics: probability distribution functions (PDFs) in the three regimes (*left*) and the cumulative distribution function (CDF) of the whole model (*right*).

Finally, the new transition probabilities are calculated as:

$$\hat{p}_{ij}^{(1)} = \frac{\sum_{t=2}^{T} P(R_t = j|x_T)\frac{p_{ij}^{(0)} P(R_{t-1}=i|x_{t-1})}{P(R_t=j|x_{t-1})}}{\sum_{t=2}^{T} P(R_{t-1} = i|x_T)}. \tag{A.10}$$

The whole procedure is repeated starting with a new parameter vector $\theta^{(1)}$. Each iteration of the algorithm generates new estimates for the smoothed inferences, as well as, new estimates $\theta^{(n)}$. Moreover, each iteration cycle increases the model likelihood function value. The two-step procedure is repeated until a (local) maximum of the likelihood function is reached.

Appendix B. Goodness-of-fit testing

The Kolmogorov-Smirnov test (K-S test) is based on measuring the difference between the empirical and theoretical cumulative distribution functions. It is used to check how well an independent and identically distributed (i.i.d.) sample fits a theoretical distribution. In the considered models, though, neither the prices themselves nor their differences or returns are i.i.d. Hence, the K-S tests cannot be applied directly to prices (or returns). Instead we use the following procedure.

First, the data is split into three (two for the 2-regime models) subsets: spikes (i.e. prices with probability $P(R_t = s|x_1, ...x_T) > 0.5$), drops (i.e. prices with probability $P(R_t = d|x_1, ..., x_T) > 0.5$) and the base regime. From formula (13) we have:

$$\varepsilon_t = \frac{X_{t+1} - (1-\beta)X_t - \alpha}{\sigma_b X_t^{\gamma}}, \tag{B.1}$$

where the ε_t's are i.i.d. Gaussian random variables. Hence, we can apply the above transformation to the base regime data and obtain three (two for the 2-regime models) i.i.d. samples: Gaussian and lognormal (or Pareto, depending on the spike regime specification) distributed, see the left panel in Figure B.10.

Combining these three (two) subsets yields a sample of independent variables with the distribution being a mixture of the lognormal (or Pareto) and Gaussian laws. The probability that a given price x_t comes from the spike distribution is equal to $P(R_t = s)$, that it comes from the drop distribution is equal to $P(R_t = d)$, while that it comes from the Gaussian law (base regime) is equal to $P(R_t = b) = 1 - P(R_t = s) - P(R_t = d)$. Therefore, the probability distribution function of the transformed sample is given by the formula:

$$f(x) = P(R_t = b)f_b(x) + P(R_t = s)f_s(x) + P(R_t = d)f_d(x), \tag{B.2}$$

where f_i is the probability distribution function of regime $i \in \{b, s, d\}$. We perform the K-S test for all subsets (base regime, spikes and drops), as well as, for the whole sample, see the right panel in Figure B.10.

References

Anderson, C.L., Davison, M. (2008). A hybrid system-econometric model for electricity spot prices: Considering spike sensitivity to forced outage distributions. IEEE Transactions on Power Systems 23(3), 927-937.

Andreasen, J., Dahlgren, M. (2006). At the flick of a switch. Energy Risk, February, 71-75.

Barlow, M. (2002). A diffusion model for electricity prices. Mathematical Finance 12, 287-298.

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

Benz, E., Trück, S. (2006). CO_2 emission allowances trading in Europe – Specifying a new class of assets. Problems and Perspectives in Management 4(3), 30-40.

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

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

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

Bunn, DW., ed. (2004) Modelling Prices in Competitive Electricity Markets. Wiley, Chichester.

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

Čižek P, Härdle W, Weron R, eds. (2005) Statistical Tools for Finance and Insurance. 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.

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

Deng, S.-J. (1998). Stochastic models of energy commodity prices and their applications: Mean-reversion with jumps and spikes. PSerc Working Paper 98-28.

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.

Geman, H., Roncoroni, A. (2006). Understanding the fine structure of electricity prices. Journal of Business 79, 1225-1261.

Hamilton, J. (1990). Analysis of time series subject to changes in regime. Journal of Econometrics 45, 39-70.

Heydari, S., Siddiqui, A. (2010). Valuing a gas-fired power plant: A comparison of ordinary linear models, regime-switching approaches, and models with stochastic volatility. Energy Economics, *forthcoming*.

Huisman, R. (2008). The influence of temperature on spike probability in day-ahead power prices. Energy Economics 30, 2697-2704.

Huisman, R., de Jong, C. (2003). Option pricing for power prices with spikes. Energy Power Risk Management 7.11, 12-16.

Huisman, R., Mahieu, R. (2003). Regime jumps in electricity prices. Energy Economics 25, 425-434.

Janczura, J., Weron, R. (2009). An empirical comparison of alternate regime-switching models for electricity spot prices. IEEE Conference Proceedings (EEM'09), DOI 10.1109/EEM.2009.5207175.

Janczura, J., Weron, R. (2010). Statistical inference for Markov regime-switching models. Working paper.

Janicki, A., Weron, A. (1994). Simulation and Chaotic Behavior of α-Stable Stochastic Processes. Marcel Dekker, New York.

Kaminski, V. (2004). Managing Energy Price Risk: The New Challenges and Solutions, 3rd ed. Risk Books, London.

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

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

Kim, C.-J. (1994). Dynamic linear models with Markov-switching. J. Econometrics 60, 1-22.

Kluge, T. (2006). Pricing swing options and other electricity derivatives. Ph.D. Thesis, University of Oxford.

Kosater, P., Mosler, K. (2006). Can Markov regime-switching models improve power-price forecasts? Evidence from German daily power prices. Applied Energy 83, 943-958.

Lucheroni, C. (2009). Stochastic models of resonating markets. Journal of Economic Interaction and Coordination, DOI 10.1007/s11403-009-0058-6.

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

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

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/.

Vasicek, O. (1977) An equilibrium characterization of the term structure. J. Financial Econometrics 5, 177-188.

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.

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