O P E R A T I O N S R E S E A R C H A N D D E C I S I O N S
No. 1 2012
DOI: 10.5277/ord120102

Carlo LUCHERONI*

A HYBRID SETARX MODEL FOR SPIKES IN TIGHT ELECTRICITY MARKETS

The paper discusses a simple looking but highly nonlinear regime-switching, self-excited threshold model for hourly electricity prices in continuous and discrete time. The regime structure of the model is linked to organizational features of the market. In continuous time, the model can include spikes without using jumps, by defining stochastic orbits. In passing from continuous time to discrete time, the stochastic orbits survive discretization and can be identified again as spikes. A calibration technique suitable for the discrete version of this model, which does not need deseasonalization or spike filtering, is developed, tested and applied to market data. The discussion of the properties of the model uses phase-space analysis, an approach uncommon in econometrics.

Keywords: *stochastic processes, time series analysis, power system economics*

1. Introduction

All physical power markets are built around the necessity of the timely and reliable physical delivery of electricity from producers to users, at socially best prices. Associated with physical markets, markets of purely financial derivatives written on the underlying electricity price processes help to manage price or quantity risks. Consequently, good mathematical models for the dynamics of electricity prices are necessary for the proper management of social welfare in the electricity industry. Each power market has its own idiosyncratic organizational characteristics, but all power markets show a common set of "stylized facts" in the dynamic features of their electricity prices, the most striking "fact" being the presence of many price spikes. These facts are very different from those found in the much more studied and much more

*School of Science and Technologies, University of Camerino, via M. delle Carceri 9, 62032 Camerino (MC), Italy, e-mail: carlo.lucheroni@unicam.it

liquid stock or bond markets. But, as usually happens in research, first models of electricity prices were developed from models of stock or bond markets [14], and in continuous time t this means, for example, geometric Brownian motion or continuous AR(1) (Ornstein–Uhlenbeck) stochastic differential equations. Unfortunately, these off-shelf models cannot reproduce typical features of power markets like seasonal random spiking. Moreover, standard stock or bond price models are designed and used to describe fundamental markets, i.e. markets that are not explicitly driven by exogenous processes like demand or volume. On the contrary, in power markets, in the presence of tight market conditions, due to capacity constraints or power grid limitations, exogenous electricity demand can strongly affect the price process. Here the word "tight" is used according to the commodity trading slang, in which it refers to a physical market condition where supply is constrained in the face of rising demand, resulting in higher prices for the commodity [37].

This paper will explore a nonlinear model of electricity prices that in certain conditions can achieve highly nonlinear effects, a model previously sketched in [30]. This model can incorporate the effect of time-varying demand on a market which can find itself in smooth or tight conditions with regard to demand levels. The model is essentially a simplified threshold autoregression, but includes a special parametric sector with unusual and useful features. Specifically, its nonlinearity will be shown to be essential in generating many of the stylized facts of electricity prices phenomenology. The model will be studied in both continuous and discrete time. In continuous time, spikes will be generated as continuous price processes, i.e. not using jumps, then in a different way from jump-diffusion models. It will be shown that the ability of the continuous time model to support spikes survives discretization. A tool not common in econometric modelling, phase-space analysis, will be used to illustrate these points. In both continuous and discrete time, it will be shown how nonlinearity can relate stochastic exogenous periodic demand to seasonal spiking and multiple mean reversion mechanisms, and how all of this can be interpreted in microeconomic terms. A calibration technique will be developed (and tested) to estimate the model using hourly data in discrete time. The focus of the paper is not on how well this model estimates the specific data set used as an example, but on how rich are the dynamics supported by the model and the range of its possibilities. In this sense, this paper is not about applied econometrics, but about investigating a new econometric tool in itself.

After the Introduction, in Section 2 some typical electricity price phenomenology will be recalled, and the link of seasonality in demand with seasonality and spiking in prices is analyzed, discussing data and market regimes. Section 3 will recall some information about discrete time and continuous time regime-switching threshold models, contrasting them with continuous time discontinuous price jump-diffusion and Lévy models. Section 4 will use phase-space analysis to show how the proposed model is able to generate spikes and how this mechanism can be related to market conditions and regimes, tight or smooth. Section 5 will consider periodic demand and

it will show how, according to the model, demand is linked to seasonality and spiking in prices, in accordance with phenomenology. Section 6 will discuss the constraints on parameters under which the model supports spikes, what happens when these constraints are violated, and in what sense the model is more general than it appears to be. Section 7 will show how to properly discretize the continuous time model. Section 8 will show how to calibrate the model in discrete time, including some calibration tests and a short analysis of results obtained using real market data. Section 9 will conclude. More details on discretizazion and calibration can be found in the two Appendices.

2. Some stylized features

In this paper, hourly price series will be considered. All the data are obtained from the AESO (Alberta, Canada) Electric System Operator website [2]. More information on this Canadian market can be found in [43], in the AESO official operating policies and procedures document [1], in the short review of Ref. [23] and in [34]. Prices are expressed in Canadian dollars C$.

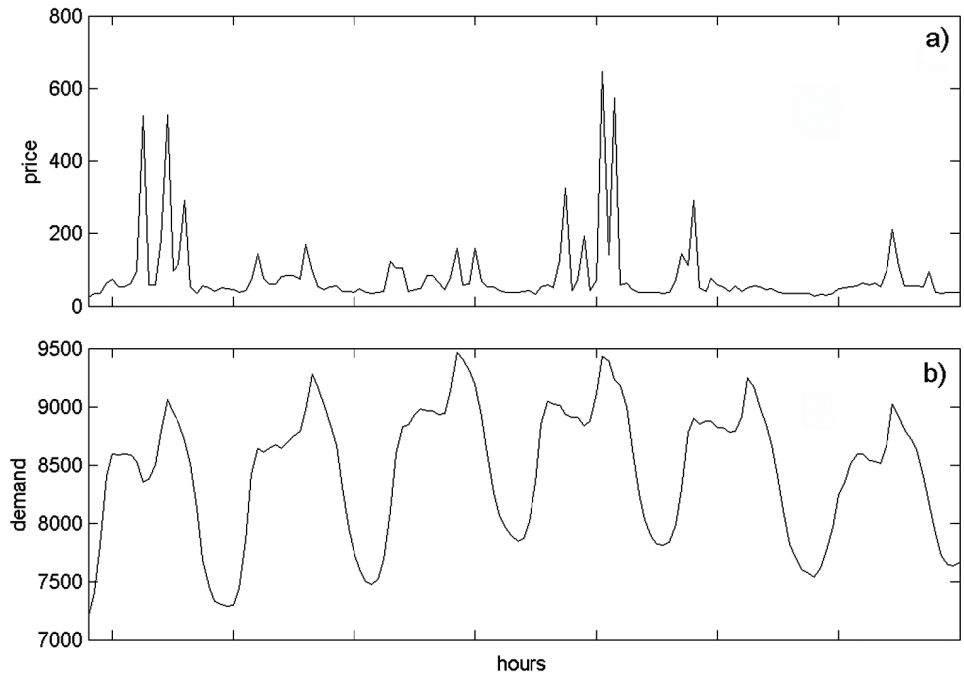

Fig. 1. Alberta power market one week from Mon Jan.08.2007 to Sun Jan.14.2007, time in hours: a) system prices (SPs) in C$ – notice some spike persistence, b) demand in MWh

The effect of demand – assumed henceforth inelastic and exogenous – on prices can be appreciated in Fig. 1, where one week of hourly prices is shown in Fig. 1a. In Figure 1b, historical demand in MWh shows day/night seasonality. Demand is highest during daylight, with a peak after lunch time. In Figure 1, electricity prices show their specific behavior in relation to demand. Prices tend to stay close to a periodically varying baseline value most of the time but sometimes rapidly increase and as rapidly revert back to their original value, tracing the shape of a spike. Noticeably, prices only spike during daylight. More precisely, sometimes they spike, sometimes they do not, but when they spike they only do so at the same time as demand crests. Thus spikes appear only occasionally but in well determined time windows. Figures 1a, b together show that at a given level of demand a spike may or may not be generated. Sometimes, before reversion, high prices can persist for a while. All considered price series show at least two reversion time scales, the time scale where spikes are involved being the shortest, the longer scale being the baseline daily seasonality. Spikes, seasonality and complex mean reversion are the most striking stylized features of power prices. The phase-space of standard financial models is too simple to allow this behaviour.

The economic origin of the spikes is not completely clear, but in the following a brief overview of theoretical (common knowledge) ideas is attempted, which relates the origin of the spikes to market tightness and behavioral factors. In a formal equilibrium approach, when demand is assumed to be inelastic, prices result from the clearing of the demand quantity q_d by the quantity–price supply curve $q_s(p)$ – assumed to be constant over time – so that the relation $q_s(p) = q_d$ sets the equilibrium price $p_{eq} = q_s^{-1}(q_d)$. Since power markets are auction pricing systems, this theoretical equilibrium approach makes sense. In a competitive environment, for a range of quantities, the supply side rationally proposes prices set at the marginal costs of production for a selection of increasingly expensive production technologies (the so-called "power stack" [13], i.e. the price-quantity curve q_s^{-1}). In aggregate, these costs also take into account production capacity constraints [9, 3], which vary in time depending on on-line and reserve generation capacity. At high levels of demand, power markets can consequently become tight because of capacity limits. Power markets transfer energy through a constraining power grid that at high demand can become congested [46], [38]. In the case of congestion, even in an abundance of capacity, power markets can then appear tight and spiky, depending on demand. At least two behavioral factors, one individual and one collective, can add volatility to this situation. Weron [44] finds that *since electricity is an essential commodity for many market participants*, in bilateral markets some demand side players individually and regularly overbid *to secure themselves a sufficient and continuous supply of power*. Moreover, since power markets are usually oligopolistic and not really competitive, anticipated tight phases can lead to collective supply side collusive behavior during these phases [34], a condition that makes prices even more spiky and volatile. Unexpected weather conditions, like extreme

temperature or low rainfall (linked to overheating of cables or hydroelectric storage shortages) can be another trigger of volatility. Thus, even using an equilibrium approach, the price-quantity curve q_s^{-1} cannot be considered to be constant over time, i.e. static. Moreover, for many markets, inelasticity of demand is an oversimplification. The main point here is that, from a broader perspective, power markets behave as if they incorporate demand threshold levels, which when exceeded take the pricing system from one regime to a different one. The level of demand selects between a normal (i.e. smooth) regime and a potentially tight regime. In the normal regime (demand off-peak hours), prices follow demand in a more or less linear way. In the potentially tight regime (demand peak hours), prices may react in a strongly nonlinear way – but not always. In fact, above a certain demand threshold, in the potentially tight regime prices can follow demand in either a linear or a nonlinear way, depending on whether capacity constraints, congestion or collusion set in.

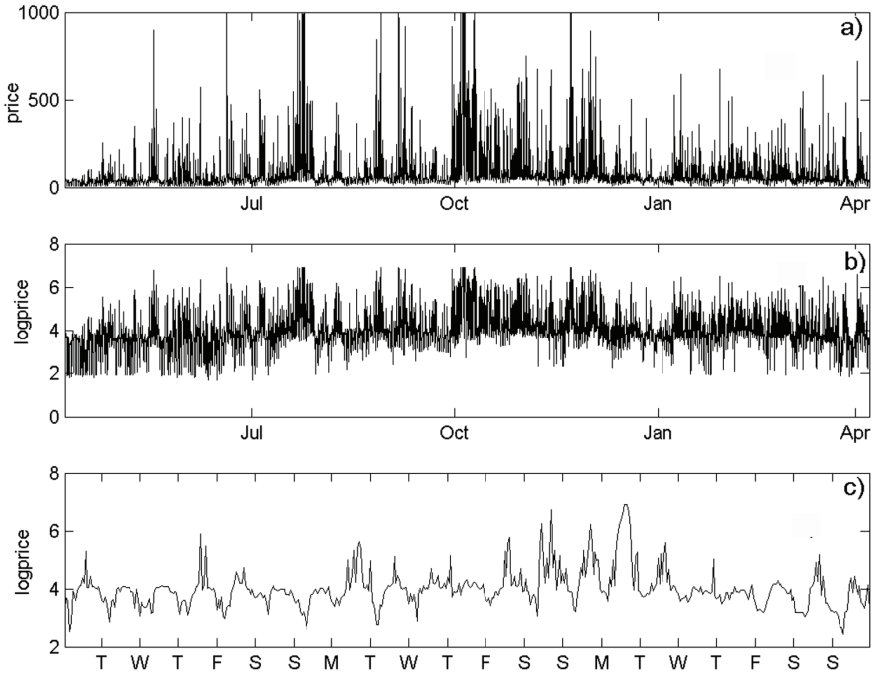

Fig. 2. Alberta power market prices: a) one year of hourly system prices p in C$, from hour 1 of Apr. 07.2006 to hour 24 of Apr. 07.2007, time ticks in months, b) hourly system logprices x, same scale, c) detail of b), 3 weeks of logprices x from hour 1 of Aug. 14 to hour 24 of Sep. 3, time ticks in week days

In Figure 2a, a 1 year time span of prices is shown, and in Fig. 2b, the logarithm of the same series is displayed. Figure 2b shows that, besides spikes, prices also undergo

antispikes, i.e. downward spikes (this is also noted in [22]). Spikes and antispikes have various heights, but three main logprice levels can be identified, i.e. a cap and a floor price level and a baseline price in between. The cap price corresponds to an institutional feature of the AESO market, where a maximum cap price cannot be exceeded. The floor price level is in principle zero (in contrast to other markets that permit negative prices). Figure 2c zooms in to show that spikes and antispikes seem to follow basically similar dynamics, but in the reverse direction and at suitably shifted times. Whereas spikes only appear during demand crests, antispikes only appear during demand troughs, and, again, only occasionally. They are probably due to the delivery at low demand levels of forward contracts, struck long in advance at far lower prices. All considered, the overall dynamics seems then to switch occasionally from a normal regime to two other different and opposite spike regimes, a potentially tight and a potentially relaxed regime. In the log plot of Fig. 2b, yearly baseline seasonality is also evident with a longer reversion time mixing with daily seasonality and spikes with a short reversion time. In this paper, only the normal and the potentially tight regimes will be modelled. Section 8 very briefly discusses how the model presented here can be elaborated further to include antispikes and the potentially relaxed regime as well.

Modelling approaches that try to reproduce spiking and seasonality in path and distributional properties only, leaving aside considerations about the economic and technical origins of spiking, are called top-down approaches. Numerical interacting--agent approaches that accurately take into account technical and institutional considerations and the individuality of economic players [7] are called bottom-up approaches. Intermediate hybrid approaches try to relate microeconomic features to stylized features using a small set of equations. In this paper, a hybrid approach will be taken.

3. Some models

In continuous time, sophisticated models have been proposed for modelling spikes taking a top-down approach. For example, Cartea and Figueroa [8] set up a first order mean reverting process with a time-dependent mean reverting level to which they attach a homogeneous Poisson process that triggers spikes. Geman and Roncoroni [15] designed a similar process with an inhomogeneous Poisson process using sharp nonlinearity to revert rising spikes to the baseline level. More general Lévy processes can also be used for this purpose (consider for example the approach by Benth [6]). All these models are based on so-called “stochastic jumps”, i.e. price discontinuities on a continuous baseline background. Calibration of this kind of jump models involves the detection of jumps by deseasonalization filtering and some spike filtering procedure. An uncommon example of a hybrid approach in continuous time not re-

quiring jumps is an equilibrium model by Barlow [4], where spikes are obtained as equilibrium prices set by a static, nonlinear and increasingly expensive price-quantity curve q_s^{-1}. The limit of such a static supply approach is that the same levels of demand generate the same price levels, which is not what happens in Fig. 1, where at a given demand level, sometimes spikes occur and at other times they do not.

In discrete time, jumps are more natural, each variation in price being potentially a jump, but in this case talking about jumps creates the problem of deciding which variation is a jump. In regime-switching state-space Markov models (RSMM) [25], which usually model daily prices, a set of S states (regimes) is defined. A different dynamics is associated with each state, and often these individual dynamics are autoregressions or simple draws from a given distribution. A higher level hidden dynamics gives the state transition from day to day. This hidden dynamics is usually written in terms of a transition matrix **M**. Such models were introduced in macroeconomics by Hamilton [17] (see also [18] for a review), using a time independent matrix **M**. Huisman et al. [21] adapted this approach to electricity price econometrics. Using a top-down approach, and pointing out that models like [15] in which jumps and baseline dynamics belong to the same regime pose difficult problems of jump identification, they overcame this difficulty by assigning baseline and spiking activity to different regimes, and carefully crafted the matrix **M** to disentangle the mean reversion times to the baseline and of spikes. This approach was further developed in [12] and [11]. It is characterized by the use of a Bayesian technique to estimate the hidden level of the dynamics [18], which has the drawback of being a bit complicated, but has the advantage of lending itself naturally to filtering, i.e. nowcasting the hidden state, and forecasting. During these developements, it was realized that **M** should be chosen to be time dependent, making it dependent on lagged prices or other exogenous time dependent variables like residual capacity, demand or temperature [33], [5] [20]. Kanamura et al. [24] were able to embed this hidden variables approach in an equilibrium model, much in the spirit of the hybrid model by Barlow. Besides RSMM, in Ref. [32] Misiorek, Trueck and Weron describe another line of modelling in discrete time, regime-switching threshold Markov models (RTMM), which use supporting regimes without hidden states (see also [45] and [22]). When the individual dynamics are autoregressions, RSMM and RTMM can be formally represented within a common frame. Consider a grid of equally spaced discrete times t_i (i = 1, ..., N), the stochastic dynamics of the price (or logprice) $X(t_i)$ and a function $u = u(E, X)$ that can depend on an external unobservable integer-valued dynamics $E(t_i)$ and on past realized values of $X(t_i)$ itself. This u is compared in value with a set $\{T_k | \, k \in K, T_1 < T_2 < ... < T_K\}$ of thresholds T_k. Depending on the result of this comparison, the system dynamics $X(t_i)$ is then associated to one of the dynamic regimes of the set $\{R_k | \, k \in K + 1\}$. If these regimes are represented by exogenously driven autoregressions (ARX) for $X(t_i)$, where the exogenous driver is $F(t_i)$, the model is called threshold ARX (TARX). If $u = u(E)$ de-

pends only on the unobservable E, the model can be called an externally excited TARX model (EETARX), and belongs to the class RSMM of state-space models. In this case, a baseline ARX regime can be associated with some values of E, and jumps can also be chosen as draws from some auxiliary random variable associated with other values of E (i.e. using AR(0) dynamics). The dynamics of E can be written in terms of **M**. If $u = u(X)$ depends only on $X(t_i)$, u is observable, the model is called a self-excited TARX (SETARX) model [39], and belongs to the class RTMM of threshold models. Under EETARX and in SETARX models, the thresholds can be assumed to be static or dynamic. Notice that **M** is not necessary in the case of RTMM but it can be derived if necessary, since it can be obtained from the form of the dynamics. A simple first order two-static-threshold ($K = 2$) three-regime example of a SETARX model is

$$\begin{cases} x_{i+1} = \phi_{R_1} x_i + \phi^0_{R_1} + f_1 + \sigma e_{i+1}, & u(x_i) \le T_1, & R = R_1 \\ x_{i+1} = \phi_{R_2} x_i + \phi^0_{R_2} + f_1 + \sigma e_{i+1}, & T_1 < u(x_i) < T_2, & R = R_2 \\ x_{i+1} = \phi_{R_3} x_i + \phi^0_{R_3} + f_1 + \sigma e_{i+1}, & T_2 \le u(x_i), & R = R_3 \end{cases} \tag{1}$$

where x_i and x_{i+1} belong to the support of the system stochastic variable X, f_i belongs to the support of F, e_i are i.i.d. draws from a distribution $P(e)$ and represent the stochastic driver, $R \in \{R_1, R_2, R_3\}$ is the regime label, and the three ϕ_R and ϕ^0_R are fixed within each regime R and are in general different from each other. The curly bracket is there to indicate to the reader that the multi-regime structure described by Eq. (1), despite its appearance, is a single SETAR(1)X equation. It could be written more compactly in terms of a regime-dependent function

$$G_R(x) = \sum_{i=1}^{3} \left(\phi_{R_i} x + \phi^0_{R_i} \right) \mathbf{1}[R = R_i] \tag{2}$$

where **1**[...] is the indicator function, equal to one when the condition in the square bracket is fulfilled and zero otherwise, as

$$x_{i+1} = G_R(x_i) + f_i + \sigma e_{i+1} \tag{3}$$

(for an example see Ref. [35]). This first order difference dynamics is globally nonlinear, even though its dynamic equations are individually linear. A second order, i.e. SETAR(2)X, dynamics can be implemented either by increasing the order of the difference, or coupling a second first order equation in an auxiliary variable y_i to Eq. (1), as for example in

$$x_{i+1} = G^x_R(x_i, y_i) \tag{4a}$$

$$y_{i+1} = G^y_R(x_i, y_i) + f_i + \sigma e_{i+1} \tag{4b}$$

(if y_i can be eliminated by substitution from the two equations, the order of the resulting equation in x_i is raised). When explicitly written in terms of coupled equations, this model is actually a vector autoregression, or SETVARX. As for an AR(2)X model, in an appropriate parametric range this model can be mean reverting (i.e. asymptotically stationary) and can sustain oscillations even in the absence of exogenous drivers. Considering that standard vector threshold regressions use the same tresholds for both G_R^x and G_R^y, a model with different x and y thresholds will be called nonstandard. In discrete time, less standard approaches to modelling electricity prices also exist, for example statistical learning techniques like support vector regressions [36] or regression trees [16]. Yet, these techniques seem to be more effective for demand modelling than for price modelling [36]. Finally, notice that in the energy economics literature, RSMM models tend to be called Markov regime-switching models, and RTMM models are called threshold models, TAR models, or threshold regime-switching models [44].

In continuous time, where autoregressions become AR diffusions, one can set up regime-switching threshold vector diffusions with the same properties of their discrete time counterparts [26]. A threshold self-excited vector diffusion (i.e., a continuous time model) that can lead to spikes and baseline sinusoidal oscillations without using jumps will be discussed now. This model consists of two coupled threshold diffusions, and it is equivalent to a suitably chosen SETVAR(1)X (or SETAR(2)X, as appropriate) in discrete time. Such a model also retains its continuous time properties in discrete time. This means that, even for the discretized model, there is no need of filtering for spikes when the model is calibrated. This model is an adaptation of a model sketched in Ref. [30]. In what follows, its regime structure will be discussed in relation to the previous discussion of tight markets, and a suitable method of calibration for it will be developed and applied to market data.

4. The McKean model: spikes as stochastic orbits

In continuous time, a model developed in mathematical neurobiology by H.P. McKean [31], discussed in Ref. [10], can be very useful in modelling spikes and can easily be related to the SETVARX models. The McKean model is based on two coupled first order equations in the variables $X(t)$ and $Y(t)$ which, in Langevin formalism, are given by

$$\varepsilon \dot{x} = g_R(x) - y \tag{5a}$$

$$\dot{y} = \gamma_b x - y + b - f(t) + \sigma(s)\xi(t) \tag{5b}$$

In Equations 5, x and y belong to the support of X and Y, $\xi(t) = \dfrac{dW}{dt}$ (where $W(t)$ is a Wiener process) is the stochastic driver, γ_b and b are the parameters, $\varepsilon > 0$ is a structural constant that determines the time scale, $\sigma(s) = \sqrt{2}s$ is the volatility set by s > 0, $f(t)$ is interpreted as the expected demand and $\sigma(s)\xi(t)$ is the random component of electricity demand. $f(t)$ is thus an exogenous driver, which can be zero or periodic. In Equation (5a), the two-threshold regime dependent function $g_R(x)$ is defined as

$$g_R(x) = \begin{cases} -\beta_L\left(x + D_L\right) - \gamma_0 D_L, & -\infty < x \le -D_L, \quad R = R_1 \\ \gamma_0 x, & -D_L < x < D_R, \quad R = R_2 \\ -\beta_R\left(x - D_R\right) + \gamma_0 D_R, & D_R \le x < \infty, \quad R = R_3 \end{cases} \tag{6}$$

where β_L, γ_0, β_R are parameters, $D_L > 0$ and $D_R > 0$ are the left (L) and right (R) thresholds, and $-D_L < D_R$. A more compact form for $g_R(x)$, analogous to Eq. (3), is

$$g_R(x) = \sum_{i=1}^{3}\left(A_{R_i}x + A_{R_i}^0\right)\mathbf{1}\left[R = R_i\right] \tag{7}$$

where, for i = 1, 2, 3, A_{R_i} and $A_{R_i}^0$ are parametric constants. In Equations (5), most of the parameters are subject to sign restrictions, i.e. $\gamma_b > 0$, $\gamma_0 > 0$, $\beta_L > 0$, $\beta_R > 0$, and to further constraints like $\gamma_0 > \gamma_b$. These constraints will be discussed in more detail in Section 6. Since the thresholds in the two equations, Eq. (5a) and Eq. (5b), are different, the model is nonstandard. Differentiating Eq. (5a) with respect to time using the chain rule

$$\dot{g} = \frac{\partial}{\partial x} g(x)\,\dot{x} \tag{8}$$

and substituting into Eq. (5b), allows one to eliminate y and rewrite the two Eqs. (5) as the single second order differential equation

$$\varepsilon\ddot{x} = \left(\frac{\partial g_R}{\partial x} - \varepsilon\right)\dot{x} + g_R(x) - \left(\gamma_b x + b\right) + f(t) - \sigma(s)\,\xi(t) \tag{9}$$

where

$$\frac{\partial g_R}{\partial x} = \begin{cases} -\beta_L, & -\infty < x \le -D_L, \quad R = R_1 \\ \gamma_0, & -D_L < x < D_R, \quad R = R_2 \\ -\beta_R, & D_R \le x < \infty, \quad R = R_3 \end{cases} \tag{10}$$

$$\frac{\partial g_R}{\partial x} = \sum_{i=1}^{3} A_{R_i} \mathbf{1}\left[R = R_i\right] \tag{11}$$

Equation 9 shows in a direct way that the McKean dynamics is essentially a second order process. $g_R(x)$ controls how $x(t)$ switches among the three AR regimes $R = R_1, R_2, R_3$ as the observable function $u(x) = x$ is compared to the two static thresholds $T_1 = -D_L$ and $T_2 = D_R$. When the restrictions on the parameters are enforced, the McKean model exhibits very interesting behavior, which can be inferred considering the two nullclines of the system. These are the two curves in the $\{x(t), y(t)\}$ phase-space such that $\dot{x} = 0$ (x nullcline y_{xn}) and $\dot{y} = 0$ (y nullcline y_{yn}) when $\sigma = 0$, i.e.

$$y_{xn}(x) = g_R(x) \tag{12a}$$

$$y_{yn}(x) = \gamma_b x + b - f(t) \tag{12b}$$

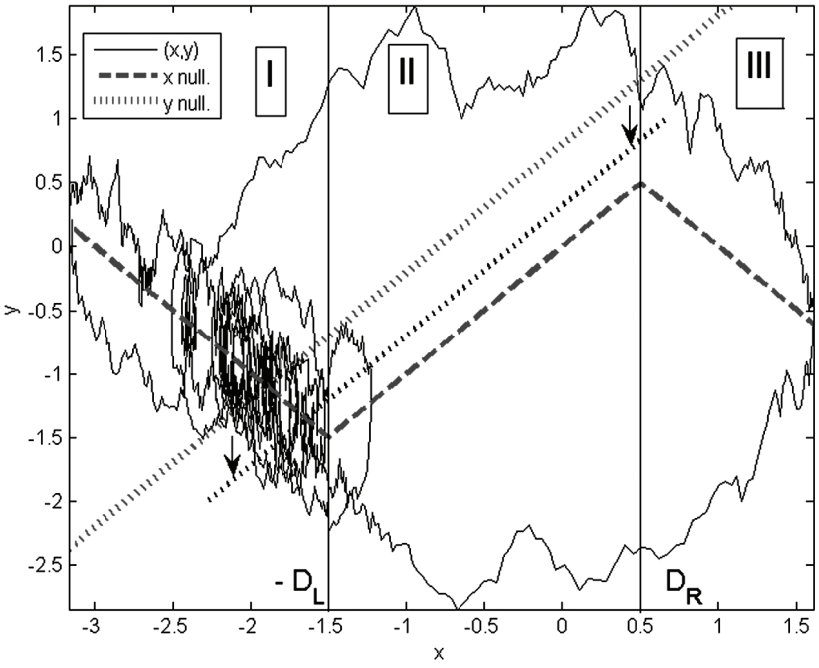

Fig. 3. McKean model phase-space, horizontal axis x and vertical axis y. A stochastic orbit, i.e. a spike, encircles most of the visible phase-space. Parameters: $\varepsilon = 0.5$, $s = 0.4$, $dt = 0.01$, $\beta_L = 1$, $\gamma_0 = 1$, $\beta_R = 1$, $y_b = 1$, $b = 0.8$, $D_L = 1.5$, $D_R = 0.5$; solid line – trajectory in phase-space dotted line – y nullcline for $f = 0$ (thinner dotted line indicated by two arrows, y nullcline shifted downward to its lowest position, i.e. for $f = -B_0$); dashed line – x nullcline. Regions are marked as I, II, III thresholds as $T_1 = -D_L$ and $T_2 = D_R$

For $f = 0$, these two curves can be seen in the phase-space plane displayed in Fig. 3 as a dashed and a dotted line, respectively. The dashed piecewise-straight line is the x nullcline that results from Eq. (12a), the dotted upward sloping line is the y nullcline that results from Eq. (12b). For the moment, keep $f = 0$ and choose the parameters of Eqs. (5) to be $\gamma_b = \beta_L = \beta_R = \gamma_0 = 1$, $-D_L = -1.5$ and $D_R = 0.5$, as in Fig. 3. Consider Eq. (6) and look at the signs in front of x in each of the three regimes. Whereas the y nullcline is straight and has a constant positive slope of value $\gamma_b = 1$, the x nullcline is segmented into three parts, which correspond to the regimes $R = R_1, R_2, R_3$. Going from left to right, the x nullcline displays the slope of -1 in R_1, a first kink due to the threshold T_1, a slope of 1 in R_2, a kink at the T_2 threshold, a slope of -1 in R_3. In this sense, the three vertical bands across the phase-space of Fig. 3 will be called regions I, II, III, since they correspond to the regimes R_1, R_2, R_3. A phase-space point $P = (x, y)$ is said to belong to the regime R_i when x takes a value in the i-th sector according to the segmentation by the thresholds. For example, if $x \in [D_R, \infty]$, then $P \in R_3$. The two nullclines cross each other at point $P_* = \{x_*, y_*\}$, where the condition expressed by taking Eq. (12a) and Eq. (12b) as a system, i.e. $\dot{x} = \dot{y} = 0$, holds. At the crossing point P_* the dynamics is quiescent, since both time derivatives are 0. P_* is a fixed point of the dynamics. In this paper, it will be assumed that the system parameters are chosen in such a way that the system is dissipative and P_* is unique and stable. P_* is then always an attractor of the dynamics. When the nullclines cross each other in region I as in Fig. 3 (which is the typical McKean setting), P_* has coordinates

$$\begin{aligned} x_*(f) &= -\frac{\beta_L D_L + \gamma_0 D_L + b - f}{\beta_L + \gamma_b} \\ y_*(f) &= \gamma_b x_*(f) + b - f \end{aligned} \tag{13}$$

In this case, since P_* is an attractor, dynamic trajectories starting close to P_* will quickly end up in P_* itself. For this reason region I will be called stable. The other two regions have different stability properties. Trajectories are also attracted by the x nullcline sector of region III (because of the nullcline's locally negative slope) but differently from what happens in region I. Attracted trajectories are quickly forced to leave, since no attractor exists in region III. For this reason, region III will be called metastable. Region II will be called unstable, since when trajectories enter this region they must cross it without stopping in it. Of course, the terms unstable and metastable are used here in a descriptive and nontechnical way since, strictly speaking, there are no fixed points except P_*. This phase-space structure is very asymmetric.

When the stochastic driver $\xi(t)$ is turned on ($f = 0$ still, but now $\sigma \neq 0$), a typical phase-space trajectory $\{x(t), y(t)\}$ is shown in Fig. 3 as a continuous orbiting line. The system spends most of its time close to where the two nullclines cross each other, the

point $P_* = P_*(f = 0)$. At a certain random time, the noise $\xi(t)$ kicks the system from region I to region II, where it cannot stop. Then the system moves on toward region III, where it can stay for a while. After having spent some time in region III, the system returns to region I. Notice that due to its two opposing trajectories through region II, the component $x(t)$ of the process has described a spike, since it has sprang up and down from its baseline value x_*. A large spike is thus a large stochastic orbit in phase-space that goes from region I to region III and back. A small spike is just a small stochastic orbit which does not reach D_R. The order of magnitude of the spike height is $T_2 - T_1$. Spike persistence appears when, during the return to region I through region II, a kick of noise with the appropriate sign pushes the trajectory back into region III. Flights from region I to region II become more likely when the nullcline y is moved closer to the second threshold T_2 (in Fig. 3, this is obtained by shifting the y nullcline in the direction of the arrows, for example changing the value of the parameter b). In this case, it is easier for the stochastic driver $\xi(t)$ to push trajectories from region I through the threshold T_1 into the unstable region II, and start a spike. In this model, spikes are not objects induced by some auxiliary process, but are endogenous to the nonlinear dynamics itself, as opposed to what happens in the jump-diffusion or state-space literature. The process $x(t)$ can now be identified as a logprice

$$x(t) = \log p(t) \tag{14}$$

In Figure 4a, a sample logprice process $x(t)$ from Eq. (5) with a time span longer than that in Fig. 3 and the same parameter values is shown. Notice that the constant mean reversion level x_*, to which the dynamics $x(t)$ always spontaneously reverts after each spike, is below the threshold $-D_L$ (inside region I, the threshold $-D_L$ is indicated by a dash-dotted line). Also notice that the dynamics naturally incorporates two reversions, the first being the spike downward movement, the second being the AR(1)-like reversion of points that never leave region I and after a noise shock turn back directly toward x_*. Figure 4b shows in which regime (I–III) the dynamics finds itself at each time. In Figure 4c, the corresponding price trajectory $p = \exp x(t)$ is shown. Price spikes with different heights and widths are present, springing from a constant baseline mean reversion level. This process is similar to the continuous time jump processes discussed in Section 3, but without jumps, i.e. the McKean process is a continuous time continuous price process. The mean reversion mechanism of the spikes is built into the equations themselves, since the McKean model incorporates a Hopf bifurcation point in its structure, which enables stochastic orbits, these occasional orbits being the spikes. To permit such spikes, of variable heights, it is very important that the diffusion is at least of second order. Notice that, by construction, the model cannot satisfy the cap price rule, while it obviously satisfies the floor price rule. The choice of the log transformation is arbitrary, and other

choices exist. But visual inspections of typical $x(t)$ patterns generated by the model (compare, for example, Fig. 4a and Fig. 2c) strongly suggest that it is the logprice transform that the model can reproduce better.

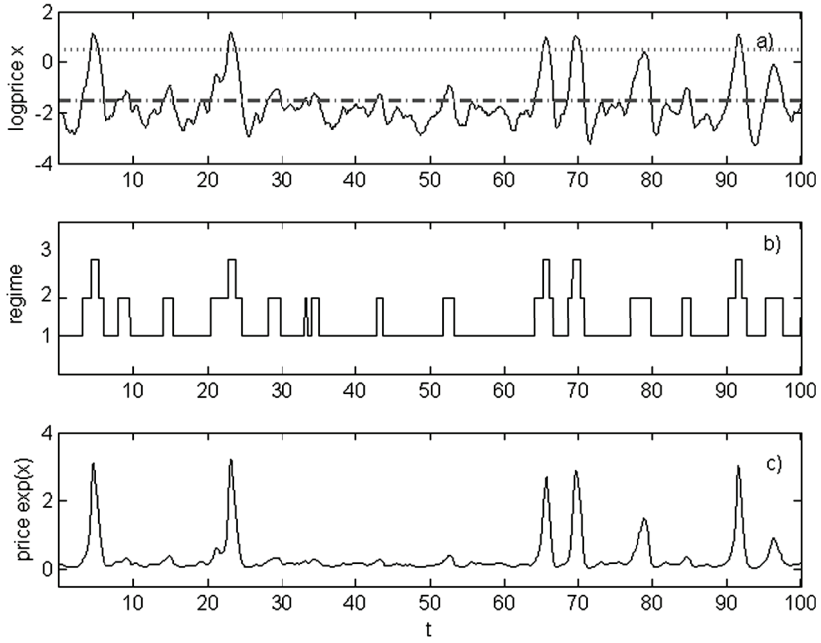

Fig. 4. McKean model for $f = 0$. Parameters: $\varepsilon = 0.5$, $s = 0.4$, $dt = 0.01$, $\beta_L = 1$, $\gamma_0 = 1$, $\beta_R = 1$, $\gamma_b = 1$, $b = 0.8$, $D_L = 1.5$, $D_R = 0.5$, t is in dt units: a) logprice $x(t)$ dynamics, dotted line D_R and dash-dotted line D_L, b) regime dynamics, c) price $p(t) = \exp x(t)$ dynamics

In Section 2, it was said that, at first sight, the level of demand selects between two regimes, a normal regime and a potentially tight regime. The first, the normal regime R_N, can thus be identified as the McKean regime R_1. The second, the potentially tight regime, can be identified collectively with the McKean regimes R_2 and R_3 as $R_{PT} = \{R_2, R_3\}$, since after the system enters region II from region I, it can either revert immediately to region I or continue to region III forming a spike. Under this interpretation, the McKean model has only two regimes, R_N and R_{PT}, in accordance with the discussion of Section 2, having assumed that the threshold T_2 is internal to the regime R_{PT}. Since the McKean model contains a threshold mechanism, it can be considered a hybrid model when it is used to describe electricity price series in terms of regimes and thresholds. More generally, the McKean model can also be useful to describe other mean reverting markets that can occasionally become tight, such as gas markets or overnight interbank rate markets.

5. The McKean model and stochastically resonating spiking

In the discussion in Section 4, the exogenous driver $f(t)$ was considered for the moment to be equal to zero. When discussing the McKean dynamics of Eqs. (5), it was said that flights through region II (i.e. the spikes) are made more likely when the y nullcline of Eq. (12b) is moved closer to the second threshold T_2 by changing the value of the parameter b. Another way to help the noise to cause spikes is obtained when the baseline f level from Eq. (5b) has a sinusoidal form, i.e.

$$f(t) = B_0 \sin \omega_0 t \tag{15}$$

In this case, $f(t)$ has a shifting effect similar to that of b, but also moves the y nullcline upward and downward in a periodic way. For small B_0, the x_* coordinate of the attractor $P_*(f)$ in region I can be periodically close to the left threshold T_1, but is never allowed to pass T_1 into region II. Consequently, the system periodically becomes more reactive to noise. If the fixed point $P_*(f)$ were allowed to pass from region I to region II, it would lose its stability, having changed its attraction properties. In this case, the dynamical system would have crossed a critical point C_p in the space of the parameters, a point that marks the location of a Hopf bifurcation. As discussed before, it is the very presence of this "external" (and never passed) critical point C_p that allows the coexistence of small oscillations and stochastically activated larger orbits (spikes). But the closer the system containing $P_*(f)$ is to C_p, the easier spikes tend to form. The combination of forcing and noise can then be interpreted as an effect of the uncertain electricity demand, and the left threshold T_1 as the soft border defining the potentially tightened market condition of the power system resulting from high demand, due to capacity constraints, grid congestion, or both. When noise is able to kick the system into region II, a spike can be fired as the trajectory tries to reach region III, but it is not necessarily fired. This feature is not present in the equilibrium model of Barlow. Moreover, in this way spike activity is mostly probable only during demand crests (daylight) and is suppressed during demand troughs (night time), an effect that in jump-diffusion models can be obtained at the price of using an inhomogeneous Poisson process that needs to be separately calibrated. Two different mean reversion scales are now present in the dynamics, one due to the fact that the day/night baseline of the $x(t)$ dynamics (the part of the dynamics without spikes) follows quasi-linearly $P_*(f) = P_*(f(t))$ in its periodicity, the other due to the fact that the spike dynamics keeps on reverting to P_*, wherever it is. The frequency of the exogenous driver can be set accordingly to the periodicity of the market, being the 24 hour seasonality most obvious for hourly data. If an exogenous driver of the type $f(t) = B_0 \sin \omega_0 t + B_1 \sin \omega_1 t$ contains extra frequencies like ω_1, time scales beyond the single-spike daylight scale can be

included and further mean reversion mechanisms can act. For example, spiking activity may be enhanced in specific high seasons (e.g. in winter, when Nordic provinces like Alberta use electric heating), while retaining the same baseline price level of other seasons. During the high season, the threshold T_1 will find itself on average closer to the demand peaks, and spikes will be more frequent – yet retain the same peak-to-base structure. As a consequence of this mechanism, the McKean model is a threshold model for reasons deeper than that of the presence of thresholds in g_R, since the critical point C_p is a threshold in itself, but with a different effect. The same combination of forcing, nonlinearity, noise and criticality was exploited as a spiking mechanism for electricity market prices in another model introduced in Ref. [27] and developed in Ref. [29] and Ref. [28], under the name of stochastically resonating spiking (SRS). A short, useful comment on the role of the time scale ε should now be made. When $\varepsilon = 0$, the second order threshold dynamics of Eq. (9) becomes the first order threshold dynamics

$$\frac{\partial g_R}{\partial x}\dot{x} = -g_R(x) + (\gamma_b x + b) - f(t) + \sigma(s)\xi(t) \tag{16}$$

with very different spiking properties. This means that ε is a stiff parameter that must take a positive value, not too close to zero. More information about the role of ε in Eqs. (5) (here tuned to the soft ε regime) can be found in Ref. [27], and a discussion about the possible uses of bifurcation theory for dynamic systems and Hopf critical points in seasonally and irregularly peaking commodity markets can be found in Ref. [29]. Ref. [28] describes how to calibrate the model introduced in Ref. [27].

An example of the effect of the SRS mechanism can be given by simulating the stochastic differential system from Eqs. (5). A typical trajectory of a forced McKean dynamics is shown in Fig. 5. The parameters used in Fig. 5 are almost the same as those used in Fig. 4 except for $b = 1$, $D_L = D_R = 1$, $B_0 = 1/2$, $\omega_0 = \pi/2$ and the Euler discretization time step $dt = 4/24 = 1/6$. Again, the process $x(t)$ is identified with the logprice, so that the price is $\exp(x)$. Figure 5a shows the simulated price trajectory, together with the sinusoidal predictable part of the demand in Fig. 5b. Each crest in demand represents daylight, each trough represents night time. This figure should be compared visually with the market data in Fig. 1. When demand pushes the system close to the left threshold $-D_L$ of region II (i.e. of the potentially tight spiking regime), indicated by a dash-dotted line, it can happen (but not necessarily) that a spike is fired. Only in this case can noise cause a spike to fire (the big arrow in Fig. 5a indicates one such event). When demand is at its trough, the probability that a spike is fired is strongly suppressed. Thus, if prices spike they do so only during daylight, and at a given level of demand a spike may or may not be generated. Moreover, even though the left threshold is crossed from left to right, in some cases the trajectory is forced back to region I very quickly without ever reaching the threshold of region III, indi-

cated by a dotted line. In Figure 5a, such a bounce appears as a would-be spike with a small height, indicated by the small arrow in Fig. 5b. Spikes appear to have a structure and not as isolated points, even though the dynamics is now discrete with 24 points per cycle.

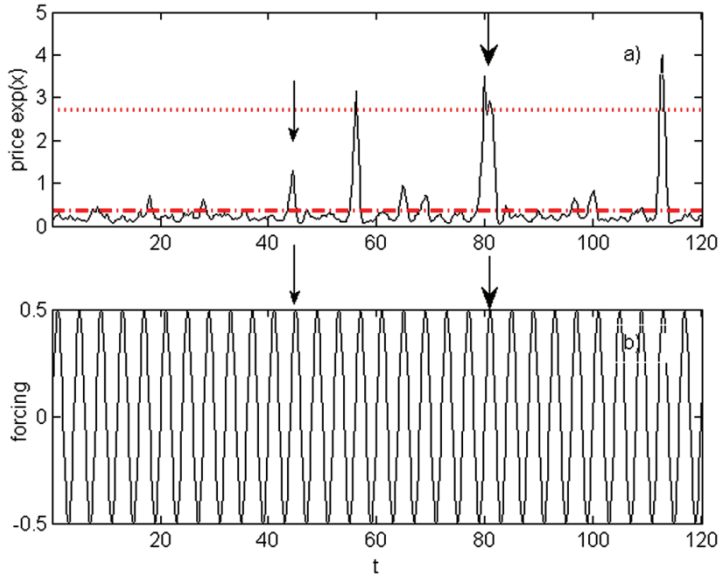

Fig. 5. Discrete McKean model for $f + 0$. Parameters: $\varepsilon = 0.5$, $s = 0.4$, $\Delta t = 1/6$, $\alpha_L = \alpha_R = 1$, $\beta_L = \beta_R = 1$, $\gamma_0 = 1$, $D_L = 1$, $D_R = 1$, $b = 1$, $\gamma_b = 1$, $B_0 = 0.5$, $\omega_0 = \pi/2$, t is in 6 hour units, i.e. one unit increment each 6 time steps (see text): a) price $\exp x(t)$ dynamics, a spike is indicated by the big arrow, a would-be spike is indicated by the small arrow, dotted line – D_R and dash-dotted line – D_L, b) forcing $f(t) = B_0 \sin \omega_0 t$, the demand crest generating the corresponding spike is indicated by a big arrow, the demand crest generating the corresponding would-be spike is indicated by a small arrow

The way in which ω_0 is chosen requires some considerations, since the choice of ω_0 has two main implications, one on the time step of the discretization and the other on the number of spikes per day. Using a simple approximation, suppose a set of values for the parameters is chosen with $f = 0$ and only then is $f = B_0 \sin \omega_0 t$ added. The chosen $\omega_0 = 2\pi/T_0$ (T_0 is the period) sets the periodicity time scale, since a full day/night cycle has to be as long as T_0. To simulate a trajectory, the number of time steps per period has to be chosen. An accurate simulation of a trajectory resulting from a nonlinear stochastic equation might require a large number of steps N_p per period to achieve a small dt, whereas the AESO market data series contains only 24 prices per period (one day). In this second case, $N_p = 24$, so that dt would turn out to be quite large. Since in the present context the McKean model is being used to model the

econometrics of electricity markets, N_p must be considered to be fixed. For a fixed N_p, ω_0 is linked to the time step Δt, because to have N_p steps per period the relation

$$\omega_0 N_p \Delta t = 2\pi \tag{17}$$

must hold. For fixed N_p, Eq. (17) shows that ω_0 is inversely proportional to Δt. Thus, an accurate simulation requires a large ω_0, i.e. a small $T_0 = N_p \Delta t$. Since ω_0 is chosen after the other parameters, the effect of ω_0 on the dynamics can be singled out by varying ω_0 itself. Under such a model, a spike is a stochastic orbit that requires some time to get back to P_0, and this time T_s is fixed when the parameters are chosen with $f = 0$. In the data, a spike starting in the morning lasts at maximum half a day. Thus, under the model, a spike must last at maximum $T_s < T_0/2 = \pi/\omega_0$. This implies that, at pre-fixed T_s, to have credible spikes, ω_0 must be bounded from above. This also limits the grid of the discretization (neither large ω_0 nor very small time steps are allowed). As a rule of thumb, one spike per day may be a good modelling choice, and this would limit ω_0 from below. This is clearly a resonance effect. Since the McKean model for $f = 0$ is a nonlinear oscillator, it has an intrinsic resonance frequency ω_s for spiking. ω_0 has to be tuned to ω_s, which is difficult to find analytically, so that some experimentation is needed. Notice that resonance effects also exist for standard linear seasonal AR(q)X models (even though they are never discussed), but here nonlinearity makes the issue particularly important. These considerations explain why the time step chosen for the simulation of Fig. 5 is so much larger than the typical dt used in the simulation of stochastic differential equations. The simulation parameters were first chosen with $f = 0$, then ω_0 was chosen to be equal to $\pi/2$, in order to have one spike per period. Consequently, $N_p = 24$ implied the choice $\Delta t = 4/24 \approx 0.17$ and $T_0 = N_p \Delta t = 2\pi/\omega_0 = 4$. Having chosen $T_0/4 = 1$ as the unit for the time axis (6 "hours", i.e. one quarter of a 24 hour period), the model was finally simulated for a time span of 30 days, i.e. for $24 \times 30 = 720$ time steps. Very importantly, this kind of simulation indicates that even long-step discrete time versions of the model preserve the features of its continuous time version.

Attaching a financial meaning to each specific parameter of Eq. (9) or Eq. (25) is not very easy, since nonlinearity mixes short term and long term reversions into one single mechanism. In any case, some of the model parameters are rather clearly related to the mean reversion velocities that the system supports, the parameters b and γ_b are linked to the probability of spiking, the quantity $D_L + D_R$ is related to the height of the spike. Finally, notice that in the discussion of the data in Section 2, the regimes correspond to the level of demand. Under this mathematical model, the demand f is exogenous but unobservable, being just a part of the mechanism, and regimes correspond to observable logprice thresholds. Yet, this does not exclude the use of an observable exogenous $f(t)$ to study the correlation of prices with external demand. The use of a substitute for the true demand has one further advantage in terms of economic inter-

pretation, since it can put to the test a power market version of the efficient market hypothesis (EMH) – all the information useful in determining future prices is contained in the past prices (and no knowledge of really exogenous variables is necessary). From the discussions in [33], [5] and [20], in power markets the EMH should reveal itself to be most certainly wrong. Power markets are not fundamental markets.

6. Constraints and generalization

The SRS mechanism based on the McKean mechanism for spiking works only if the slopes of the nullclines have the appropriate signs and the critical point is never passed. This implies that the parameters are subject to constraints. Subject to sinusoidal forcing, the y nullcline moves up and down in a band limited by the two extremes $\pm B_0$ that $f(t)$ can reach (its lowest position is indicated in Fig. 3 by the thinner dotted line). For a given choice of parameters, when $x = -D_L$, the y nullcline $y_{yn}(x)$ attains its minimum at the value $y_{yn}^{\min}(-D_L) = y_b(-D_L) + b - B_0$. If the relation $y_b \geq \gamma_0$ is assumed, a sufficient condition for the attractor point P_* to remain in region I (so that the critical point is not passed) is that at $x = -D_L$ the y nullcline remains above the x nullcline. That is, $y_{yn}^{\min}(-D_L) \geq y_{xn}(-D_L) = -\gamma_0 D_L$. Combining all these constraints, the SRS mechanism works if

$$\begin{cases} -D_L\gamma_b + D_L\gamma_0 + b - B_0 \geq 0 \\ \gamma_b - \gamma_0 \geq 0 \\ \beta_L, \beta_R, \gamma_0 \geq 0 \end{cases} \tag{18}$$

(which implicitly implies $\gamma_b \geq 0$). It should be noted that if in the continuous time model given by Eq. (9) the parameters are not constrained, the model encompasses the harmonic damped forced oscillator, i.e. second order linear diffusion, as a special case. This occurs because, when $\gamma_0 < 0$ and $\beta_L = \beta_R = -\gamma_0$, $g_R(x)$ is equal to $\gamma_0 x$ in all three regimes, which when substituted into Eq. (9) gives the linear oscillator equation

$$\varepsilon\ddot{x} = (\gamma_0 - \varepsilon)\dot{x} + \gamma_0 x - (\gamma_b x + b) + f(t) - \sigma(s)\xi(t) \tag{19}$$

For parameter values between the McKean spiking model and the harmonic oscillator, i.e. when $\gamma_0 < 0$, $\beta_L > 0$, $\beta_R > 0$ but $\beta_L \neq -\gamma_0$ and $\beta_R \neq -\gamma_0$, the model behaves like a quasi-linear mean reverting oscillator, since the x nullcline is almost straight, and the phase-space structure is now almost symmetric. Even this form can be useful because it incorporates threshold effects while not requiring the McKean mechanism to be at work.

Thus, not to lose generality, in continuous time it is better to consider the (constrained) McKean model as a restriction of the more general model given by Eq. (9) (or Eqs. (5)) whose parameters are not constrained. A second possible restriction of the model is the harmonic damped oscillator. A third useful restriction is the quasi-linear (or weakly nonlinear) mean reverting model.

7. Discretization

When an econometric calibration method has to be implemented in the model given by Eqs. (5), a clear problem appears. The process $X(t)$ can be directly estimated using the logarithm of price data, but it might not be immediately clear on what basis the auxiliary process $Y(t)$ could be estimated. However, note that Eqs. (5) or Eq. (9) can be further (and trivially) represented by the system

$$\dot{x} = z \tag{20a}$$

$$\varepsilon \dot{z} = \left(\frac{\partial g_R}{\partial x} - \varepsilon \right) z + g_R(x) - \left(\gamma_b x + b \right) + f(t) - \sigma(s)\xi(t) \tag{20b}$$

where z belongs to the support of a new auxiliary variable Z. In this case, the chain rule Eq. (8) is not necessary to go from Eqs. (20) to the second order Equation (9). Since

$$z = \dot{x} = \frac{\dot{p}}{p} \tag{21}$$

is clearly an instantaneous logreturn, after a discretization indexed by n with time step Δt

$$\dot{x}(t) \to \frac{x_{n+1} - x_n}{\Delta t} \tag{22}$$

the variable z in Eq. (21) becomes the logreturn intensity variable

$$z_n = \frac{x_{n+1} - x_n}{\Delta t} \tag{23}$$

and all the data used for a sample series $\hat{x}_i$ can be used to build the auxiliary sample series $\hat{z}_i$. The Euler discrete time version of Eqs. (20) becomes

$$\frac{x_{n+1} - x_n}{\Delta t} = z_n \tag{24a}$$

$$\frac{\varepsilon\left(z_{n+1}-z_n\right)}{\Delta t}=\left(\frac{\partial g_R}{\partial x}-\varepsilon\right)z_n+\left(g_R\left(x_n\right)-\left(\gamma_b x_n+b\right)+f\right)-\frac{\sigma(s)}{\sqrt{\Delta t}}\eta_{n+1} \tag{24b}$$

where the η_n are i.i.d. variables with a $N(0, 1)$ distribution, $g_R(x)$ is given by Eq. (6) and $\frac{\partial g_R}{\partial x}$ is given by Eq. (10). Direct use of the discretization formula of Eq. (22) in Eq. (9), or substitution of Eq. (24a) into Eq. (24b), give the Euler discrete time version of Eq. (9)

$$\begin{aligned}\frac{\varepsilon\left(x_{n+2}-2x_{n+1}+x_n\right)}{(\Delta t)^2}=&\frac{\partial g_R}{\partial x}\frac{x_{n+1}-x_n}{\Delta t}-\frac{\varepsilon\left(x_{n+1}-x_n\right)}{\Delta t}\\&+g\left(x_n\right)-\left(\gamma_b x_n+b\right)+f\left(t_n\right)-\frac{\sigma(s)\eta_n}{\sqrt{\Delta t}}\end{aligned} \tag{25}$$

As briefly sketched in Appendix A, two other interesting discretizations (including the discretization of Eqs. (5)) are still possible but they will not be discussed any further. In Appendix B, it will be shown that under self-consistent assumptions the unconstrained calibration of Eq. (25) and Eqs. (24) (as opposed to unconstrained calibration of the discrete version of Eqs. (5)), is not sensitive to the value of ε. The discrete time version of the general process represented by Eq. (9) is then ready to be used in electricity finance either as the single second order autoregression (Eq. (25) (a SETARX model), or equivalently as the two coupled first order regressions (Eqs. (24)) (a SETVARX model). Both forms include three possible restrictions, or modalities. The simplest one is an AR(2)X (or a vector AR(1)X) model, equivalent to the harmonic oscillator. Another one is a quasi-linear TARX (or TVARX) model. A third modality is obtained when the parameters are subject to the constraints given in Eq. (18). This third, strongly nonlinear restriction supports spiking and is the discrete time version of the McKean model. Since the discrete time version of the general unconstrained process of Eq. (9) encompasses such a rich phenomenology, it should be rewarded with a name. It will be called S-SETARX, i.e. spiking SETARX. It is interesting to notice that when Tong introduced threshold models in Ref. [40], he was interested mainly in their potential to support strongly nonlinear behavior (as widely discussed in his book [39]) but in the end only the milder and quasi-linear aspects of these models entered the mainstream econometric literature (mainly due to the work of Tsay who implicitly worked on standard models [41], [42]). This notwithstanding, since in the end the S-SETARX model is just a two-threshold SETARX (even a bit less flexible, because the model is nonstandard and the y nullcline is not even broken), it is rather strange that it has never been taken into consideration in electricity finance before.

8. Calibration

Besides ε, fixed for the moment as a structural constant, and ω_0, the S-SETARX model depends on the following set of six parameters

$$\Psi = \{\beta_L, \gamma_0, \beta_R, \gamma_b, b, B_0\} = \{\psi_1, \psi_2, \psi_3, \psi_4, \psi_5, \psi_6\} \tag{26}$$

plus σ, plus the two thresholds

$$\Theta = \{-D_L, D_R\} = (T_1, T_2) \tag{27}$$

i.e. nine parameters in all. Some of these parameters can be subject to constraints like those in Eq. (18). The simplest way to calibrate a two-threshold SETAR model is to follow the approach of Ref. [19], using sequential loglikelihood L maximization. This procedure consists of 1) choosing an arbitrary pair of thresholds, 2) partitioning the data into regimes using the chosen thresholds, 3) inside each regime estimating the model's parameters by loglikelihood maximization independently of the other regimes, 4) collecting these estimates and using them to build a common estimate for the volatility by means of the residuals, and 5) finding some way to improve the estimates of the pair of thresholds until a global minimum of the volatility (and thus a maximum of the loglikelihood L) is reached. At this minimum volatility, the maximum likelihood estimators of the model's parameters and thresholds are obtained. As a SETAR system, two features are peculiar to the S-SETARX model. First, being a nonstandard model, each regime has a specific Eq. (5a) but Eq. (5b) is the same for all regimes. This case was not included in Ref. [19] and this complicates the estimation process, since the three regimes cannot be estimated independently. Second, because of the intrinsic scarcity of spike data with respect to baseline data, the regime that contains spikes can lead to large estimation errors that are reflected globally in the estimate of σ, so that adaptive methods for sequentially estimating thresholds might be inefficient in comparison to simpler grid searches. In Appendix B, it is shown how calibration can be achieved in both constrained and unconstrained cases, and how the problem of regime coupling can be resolved. In the case of unconstrained calibration, if Ψ takes the form of the row vector $\boldsymbol{\psi}$, it is possible to numerically build a matrix $\mathbf{C}$ and a row vector $\mathbf{V}$ such that an estimate of $\boldsymbol{\psi}$ is given by the inversion of

$$\mathbf{C}\boldsymbol{\psi}' = -\boldsymbol{\varepsilon}\mathbf{V}'$$

where the apostrophe means transpose.

Unconstrained calibration is the first kind of calibration that should be attempted, since it can explore all the modalities of the S-SETARX model. Depending on the outcome, the signs of the estimated parameters can give information on whether the

data can be best represented by a spiking system, or a harmonic oscillator (i.e. a simple AR(2)X), or a quasi-linear oscillator. Then, the constrained calibration described in Appendix B can be tried, to see whether imposing constraints can force the estimate to some local minimum (the system is nonlinear), to collect more information.

Table 1. Calibration on synthetic data

$-\hat{D}_L = -1, \quad \hat{D}_R = 1$						
$\hat{\beta}_L$	$\hat{\gamma}_0$	$\hat{\beta}_R$	$\hat{\gamma}_b$	$\hat{b}$	$\hat{B}_0$	$\hat{\sigma}$
1.0454	1.0406	1.0458	1.0113	0.9970	0.4768	0.8934

The unconstrained calibration procedure described in Appendix B can be tested by generating a series of synthetic data with the same parameters as those used for Fig. 5 (where $\sigma = 0.8944$), but with a length corresponding to one year (8759 values). Implementation of the procedure for $N_p = 24$, the external values $\varepsilon = 0.5$, $\omega_0 = \pi/2$ (and consequently $\Delta t = 1/6$) on a grid of values $-D_L \in [-1.8, -0.2]$ with step size 0.1 and $D_R \in [0.2, 1.8]$ with step size 0.1 (which includes (–1, 1)) leads to the pair of estimates $(-\hat{D}_L = -1, \ \hat{D}_R = 1)$.

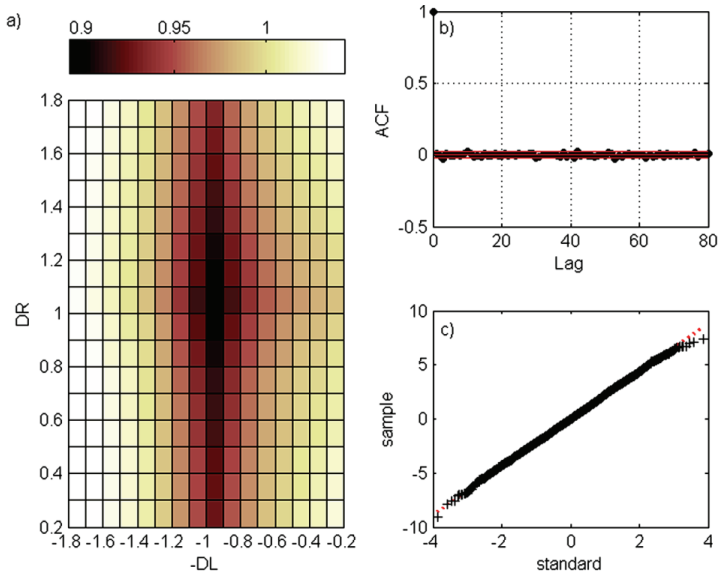

Fig. 6. Calibration on synthetic data: a) volatility surface for $(-D_L, D_R)$ pairs. The gray scale indicates the level of estimated volatility for each threshold pair $(-D_L, D_R)$ considered, darker shadows correspond to lower values of volatility. The best estimate of $(-D_L, D_R)$ corresponds to the surface minimum, b) autocorrelation function (ACF) of the residuals, c) Q–Q plot of the residuals

A plot of the volatility surface corresponding to the chosen range of threshold pairs is shown in Fig. 6a, where darker shadows correspond to lower estimates of volatility. The localization of the left threshold at $-\hat{D}_L$ is sharp, the localization of the right threshold $\hat{D}_R$ is smoother. Since D_R is below the peak of the spikes, this signals that this method is not too sensitive to the position of the threshold defining spikes. This makes $\hat{D}_R$ an interesting possible data-driven estimate of a threshold to discriminate spikes from non-spikes. The estimated values of the other parameters are reported in Table 1. The estimated maximum loglikelihood is –19 420. Figure 5b illustrates the autocorrelation function (ACF) for the residuals of the series, which shows no residual autocorrelation. Figure 5c gives a normal Q–Q plot for the residuals. Since all the data lie very close to the diagonal line, the distribution is normal. A Jarque–Bera test of the null hypothesis that the sample of residuals comes from a normal distribution returns that normality is not rejected with a *p*-value of 0.5. The estimated mean of this normal distribution is $\mu_d = 10^{-15}$. The calibration procedure seems to work well.

The logprice series of 8760 real data for the year between Apr. 7.2006 and Apr. 7.2007 is shown in Fig. 2b. To estimate the parameters of model using these data, the fixed values of $\varepsilon = 0.5$, $\omega_0 = \pi/2$ (thus $\Delta t = 1/6$), and a grid with $-D_L \in [2.6, 4.4]$ with the step size 0.1 and $D_R \in [5, 6.6]$ with the step size 0.1 were chosen (previously, a preliminary search on a larger grid had been made). The pair of thresholds which minimize volatility and maximize the loglikelihood is ($-\hat{D}_L = 4.1$, $\hat{D}_R = 6.3$), with an estimated optimizing phase shift index (see Appendix B) of $\hat{l}_\phi = 11$. A plot of the volatility surface corresponding to the chosen threshold pairs is shown in Fig. 7a. The estimates of the parameters are reported in Table 2.

Table 2. Calibration using AESO market data (Apr. 7.2006–Apr. 7.2007)

$-\hat{D}_L = -4.1, \quad \hat{D}_R = 6.3, \quad \hat{l}_\phi = 11$						
$\hat{\beta}_L$	$\hat{\gamma}_0$	$\hat{\beta}_R$	$\hat{\gamma}_b$	$\hat{b}$	$\hat{B}_0$	$\hat{\sigma}$
3.5496	–4.0098	3.0303	0.2114	–16.6987	–3.5486	3.7803

The estimated maximum loglikelihood is –31916. The ACF of the residuals is shown in Fig. 7b. Some correlation at daily lags of multiples of 24 hours remains. The Q–Q plot in Fig. 7c shows that the distribution of the residuals is not normal, especially in the tails. The Jarque–Bera test rejects normality. Figure 8 shows how the residuals are distributed over time with relation to the original time series. In Figure 8a, *b* prices and logprices are shown for a portion of the series where there is a high density of spikes. Calibration is at its best when errors at the rising front of a spike are low (and similarly when the spike drops down). If this is not the case, the

system dynamics is not reactive enough, i.e. one of the reversion times is not sufficiently short. In the terminology of Fourier analysis, the system does not support frequencies high enough to be accurate around sharp features of the signal. Small errors in periods in which only baseline activity is present indicate that, on the contrary, the system can follow smooth signals, and, in this case, it does not need ether reactivity or high frequencies.

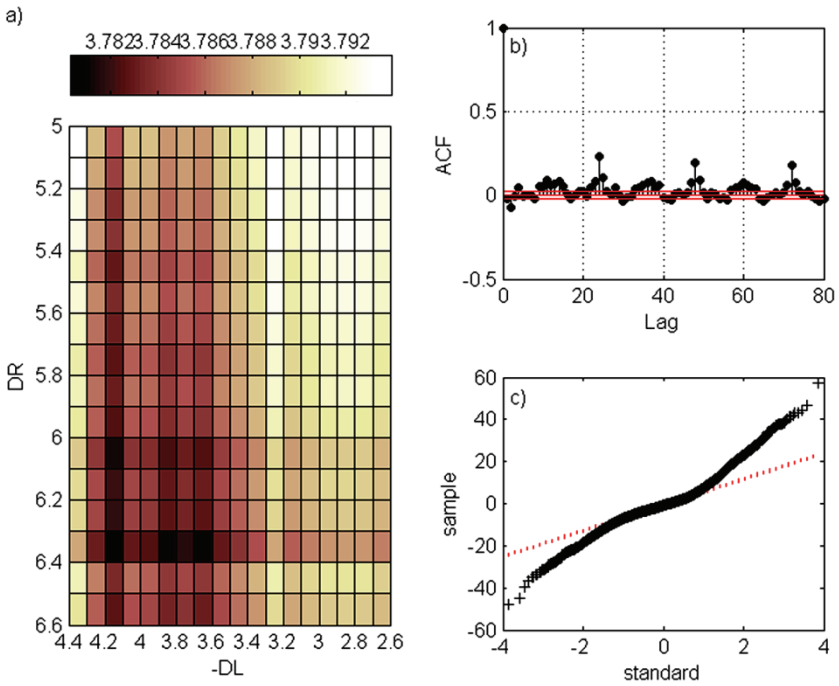

Fig. 7. Calibration on AESO market data (Apr. 7.2006–Apr. 7.2007):
a) volatility surface for $(-D_L, D_R)$ pairs. The gray scale indicates the level of estimated volatility for each threshold pair $(-D_L, D_R)$ considered, darker shadows correspond to lower values of volatility. The best estimate of $(-D_L, D_R)$ corresponds to the surface minimum, b) autocorrelation function (ACF) of the residuals, c) Q–Q plot of the residuals

Figure 8c shows the estimates of the residuals based on the estimates of the parameters in Eq. (B.10). Errors are rather large at the fronts of spikes, and smaller when there is just baseline activity, which is typical of AR(q) systems with low q (thus, with a small range of frequencies). This poor quality of fit can be explained by looking at the sign that the procedure finds for $\hat{\gamma}_0$, a negative sign. This implies that the SETAR(2)X system fits the data using the form of a quasi-linear oscillator, and not in the highly nonlinear spiky form resulting from a McKean model. Even though spikes are present in the data, it seems that the system finds a fit with smooth oscillations to be more

likely (high L) than a McKean model. Calibration constrained to the McKean modality does not detect anything interesting. The smooth quasi-linear form chosen is not completely satisfactory and, besides, from Fig. 8c it can also be seen that autocorrelation is not completely removed from the residuals. To better understand this point, some AR(q)X models were fitted to the data (no moving average terms added, to remain in the same class of SETAR(2)X, and using a sine function as an external driver). Increasing q improves the estimates, since higher frequencies are introduced, and a model identification procedure finds $q = 8$ to be the best value. The ACF for $q = 8$ is no better than that for the nonlinear model, which exploits nonlinearity to achieve the same effect with fewer lags.

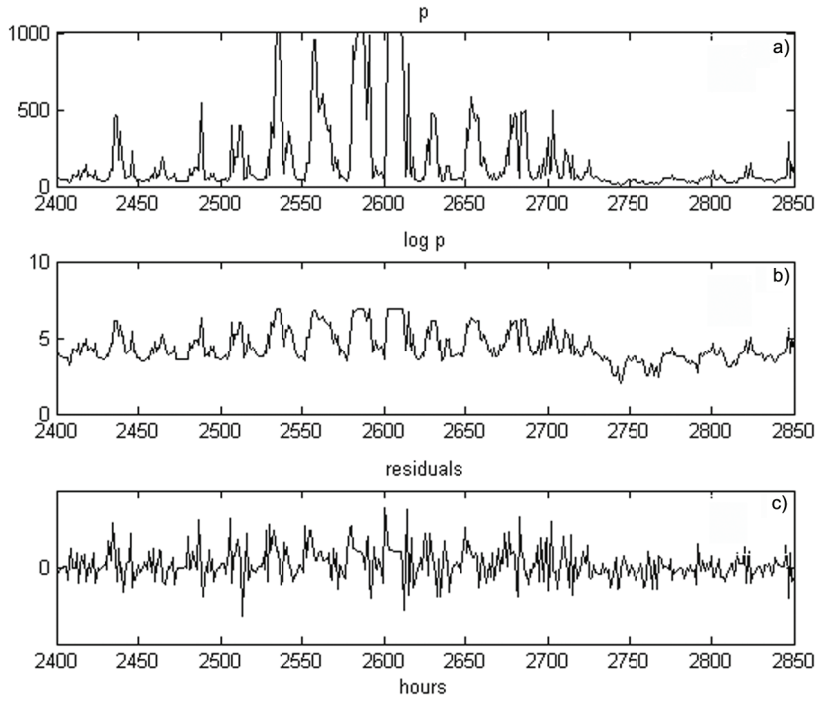

Fig. 8. Residuals. Abscissas in hours: a) prices, b) logprices, c) residuals (see text)

The reason why the calibration chooses the thresholds reported in Table 2 can be understood by taking a look at the logprice Poincare plot in Fig. 9. A Poincare plot is a scatterplot where data points from a series w_n are displayed on the plane as (w_{n+1}, w_n) pairs. In Figure 9, $w_n = \log p_n$. In the AESO series, each value $\log p_n$ (read on the abscissa) corresponds to a relatively small interval of logprices $\log p_{n+1}$ at the next moment (read on the ordinate), which is level dependent. For example, very small logprices, below 2, can be followed only by very small logprices. The eye detects at least

three big regions in Fig. 9 (in fact there are five). Logprices in different regions are subject to different dynamics (and followed by logprice ranges that depend on the starting region), which reinforces the "normal and potentially tight market" hypothesis discussed in Section 2, which implies that these thresholds are important for the AESO data. Three (five) regimes correspond to two (four) thresholds. When the two--threshold S-SETARX is fitted to these data, the calibration procedure, even though not switching to the spiking sector, exploits the flexibility of the model, adjusting $-D_L$ and D_R as in Fig. 9, i.e. incorporating the largest "square region" between the two thresholds, and leaving the rest of the data to the two external segments of the x nullcline.

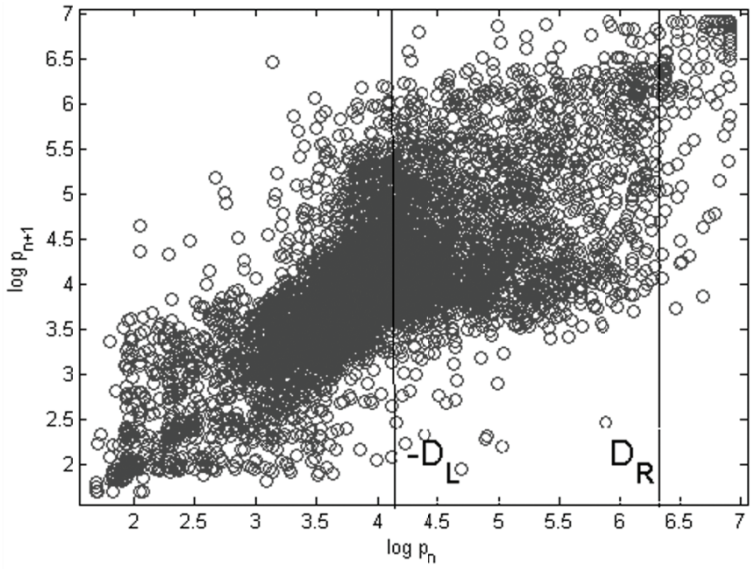

Fig. 9. Poincaré plot of logprices: on the abscissa $x_n = \log p_n$, on the ordinate $\log p_n$

Looking again at Fig. 7, besides the lowest-volatility (darkest) square, for $D_R = 6.3$ the calibration finds two possible lower thresholds at $-D_L = 3.8$ and at $-D_L = 3.6$, and for $D_R = 6.0$ three lower thresholds at $-D_L = 4.1$, 3.8, 3.6. From Figure 9, this is explained by the fact that the largest square region is not sharply defined, and there are different quasi-equivalent ways of framing it. The reason why the model chooses to lock itself into the quasi-linear modality may be explained as follows. In Section 2, it was stated that the real logprice series shows spikes (only at the same time as demand crests) and antispikes (only at the same time as demand troughs). Thus in such series there is a sort of symmetry between the data above and below the baseline reversion level. The McKean model in its spiking form (as all other models for logprices that include spikes, but do not include antispikes) is essentially asymmetric with respect to

this level. This implies that the weight of the spike at crests is not balanced by the weight of the antispike at the appropriate time lag. The best solution that the calibration procedure finds to fit the data is to choose the most symmetric version of the model, i.e. its quasi-linear oscillator form. A way to overcome this problem of asymmetry for the spiking McKean model would be to extend the McKean model by defining an antispike sector, adding a metastable and an unstable region at the left hand side of the left threshold. Such an extended model was the model sketched in Ref. [30], for which no calibration has been tried or developed. Since the model of Ref. [30] includes antispikes, an extension to that model of the calibration procedure presented here could be useful in understanding whether the SETAR(2)X and the McKean model fail in accurately explaining market data just because of a lack of an antispike sector. Such a developement is left to further work. Another development would be a comparison of calibrations obtained for a selection of different ω_0, to investigate the importance of the resonance effects discussed in Section 5. Yet, as stated in Section 1, an accurate description of the econometrics of the AESO data is not the main purpose of this paper, which instead is a discussion of the interesting features and possibilities of the S-SETARX model.

9. Conclusions

The S-SETARX model is a discrete time nonlinear econometric model, born in continuous time, with a rich phenomenology, an intriguing dynamics, and encompassing some stimulating mathematical subtleties. It has three main modalities, one of which heavily exploits its non-linearity and supports spikes in a very natural way. Being a threshold model, it can be used to analyze data sets where spikes are present and thresholds are suspected, and in this way it can be useful even if the spike generating mechanism hidden in the data is different from that supported by the model itself. In this paper, the model was discussed focussing attention on the case in which an exogenous sinusoidal driver is present, but, as discussed in Section 4, the same model could also be very useful without the sinusoidal demand term, for example when analyzing daily or monthly averaged data or data from spiky time series that do not depend on seasonality. In conclusion, the S-SETARX model, endowed with 1) its microeconomic interpretation of the mathematically defined normal and potentially tight regimes, 2) its extension of the meaning of the world “threshold” in modelling spikes (intended both as a kink in a nullcline and as a bifurcation point leading to stochastic orbits) and 3) its associated constrained and unconstrained calibration procedure, is certainly rich enough in phenomenology to be included in the set of techniques that can be used to model electricity prices.

Acknowledgements

The author wishes to thank S. Sapio and R. Weron for two interesting discussions about TAR models, C. Pacati for many discussions about the SRS mechanism, and D. Bunn for a discussion about the literature on hybrid models.

Appendix A. Discretizations

In continuous time, elimination of y from the McKean Eqs. (5), or elimination of z from Eqs. (20), gives the second order Eq. (9). Direct use of the discretization described in Eq. (22) on the McKean Eqs. (5) gives the first order system

$$\frac{\varepsilon\left(x_{n+1}-x_n\right)}{\Delta t}=g\left(x_n\right)-y_n \tag{A.1a}$$

$$\frac{y_{n+1}-y_n}{\Delta t}=\gamma_b x_n-y_n+b-f\left(t_n\right)+\frac{\sigma(s)\eta_n}{\sqrt{\Delta t}} \tag{A.1b}$$

Elimination of y_n from Eqs. (A.1) gives the second order equation

$$\begin{aligned}\frac{\varepsilon\left(x_{n+2}-2x_{n+1}+x_n\right)}{\left(\Delta t\right)^2}=&\frac{g\left(x_{n+1}\right)-g\left(x_n\right)}{\Delta t}-\frac{\varepsilon\left(x_{n+1}-x_n\right)}{\Delta t}\\&+g\left(x_n\right)-\left(\gamma_b x_n+b\right)+f\left(t_n\right)-\frac{\sigma(s)\eta_n}{\sqrt{\Delta t}}\end{aligned} \tag{A.2}$$

which is not Eq. (25), the discretization of the second order Eq. (9), since in general

$$\frac{g\left(x_{n+1}\right)-g\left(x_n\right)}{\Delta t}\neq\frac{\partial g}{\partial x}\left(x_n\right)\frac{\left(x_{n+1}-x_n\right)}{\Delta t} \tag{A.3}$$

because

$$\frac{\partial g}{\partial x}\left(x_n\right)\neq\frac{g\left(x_{n+1}\right)-g\left(x_n\right)}{x_{n+1}-x_n} \tag{A.4}$$

However, the continuous limit of Eq. (A.4) is Eq. (9). This could seem strange but it is due to the fact that, in continuous time, to go from Eqs. (5) to Eq. (9) the use of the chain rule

$$\dot{g} = \frac{\partial}{\partial x} g(x)\,\dot{x}$$

was made, whereas the chain rule was not involved when going from Eqs. (20) to Eq. (9), and the chain rule does not hold for arbitrary large Δt in discrete time, but only at the $\Delta t \to dt$ limit. Thus, Eq. (25) and its associate Eqs. (24) (which are called in this paper the S-SETARX model) are dynamic maps different from the maps of Eq. (A.2) and its associate Eqs. (A.1), with different dynamic properties at arbitrarily large time steps. All these four dynamic maps have the same continuous time limit, Eq. (9).

Appendix B. Details of the calibration

Calibration will be carried out on the $\{x_n, z_n\}$ SETVARX system of Eqs. (24), Eq. (6) and Eq. (10). The likelihood for a sequence of N sample data can be written as the product

$$H = \prod_{n=2}^{N} p(n+1|n) \tag{B.1}$$

where $p(n + 1|n)$ is the conditional probability density of $\{\hat{x}_{n+1}, \hat{z}_{n+1}\}$ given $\{\hat{x}_n, \hat{z}_n\}$, where the initial marginal density $p(1)$ is dropped for simplicity. $p(n + 1|n)$ can be factorized into x and z components as $p(n + 1|n) = p_x(n + 1|n)p_z(n + 1|n)$, where $p_x(n + 1|n) = 1$, since Eq. (20a) contains no noise. The noise term $\sigma\xi$ in Eq. (20b) has a Gaussian distribution $N(0, \sigma^2)$, so that it is more convenient to work with the logarithm of H, i.e. with the loglikelihood

$$L = L_0 + L_\sigma \sum_{n=2}^{n} \left(\Omega_R(n+1, n)\right)^2 \tag{B.2}$$

where

$$L_0 = -\frac{N-1}{2} \ln \frac{2\pi\sigma^2}{\Delta t} \tag{B.3}$$

$$L_\sigma = -\frac{\Delta t}{2\sigma^2} \tag{B.4}$$

and

$$\Omega_R(n+1, n) = \sum_{i=1}^{3} \Omega_{R_i}(n+1, n)\,\mathbf{1}[\hat{x}_n \in R_i] \tag{B.5}$$

In Equation (B.5)

$$\Omega_{R_i}(n+1,n)=\frac{\varepsilon(\hat{z}_{n+1}-\hat{z}_n)}{\Delta t}-\left(\frac{\partial g_{R_i}}{\partial \hat{x}_n}-\varepsilon\right)\hat{z}_n-g_{R_i}(\hat{x}_n)+(\gamma_b\hat{x}_n+b)-f(t_n) \quad \text{(B.6)}$$

Because of the indicator function, at each n the coordinate x_n of point P_n is examined, and P_n is ascribed to the appropriate regime R_i, so that the proper component $g_{Ri}(x)$ of g_R is selected. After defining

$$h_n=h(n+1,n;\Delta t)=\frac{\hat{z}_{n+1}}{\Delta t}+\hat{z}_n\left(1-\frac{1}{\Delta t}\right) \quad \text{(B.7)}$$

Equation (B.6) can be written as

$$\Omega_{R_i}(n+1,n)=\varepsilon h(n+1,n;\Delta t)-\frac{\partial g_{R_i}}{\partial \hat{x}_n}\hat{z}_n-g_{R_i}(\hat{x}_n)+(\gamma_b\hat{x}_n+b)-B_0\sin\omega_0 t_n \quad \text{(B.8)}$$

or, for

$$A^0(n+1,n)=\varepsilon h(n+1,n;\Delta t) \quad \text{(B.9)}$$

in a form more explicitly linear in the system parameters Ψ as

$$\Omega_{R_i}(n+1,n)=A^0(n+1,n)+\sum_{i=1}^{6}A_{R_i}^{j}(n)\psi_j \quad \text{(B.10)}$$

Equation (B.10) defines the data matrix $A_R(n)$ (with 6 × 3 entries $A_{\text{regime}}^{\text{parameter}}(n) = A_{R_i}^{j}(n)$, j = 1, ..., 6, i = 1, 2, 3) where each j is associated with one of the six parameters in Ψ, each i with one of the three regimes. In terms of the parameter labels of Eq. (6)

$$\begin{aligned}\Omega_{R_1}(n+1,n)=A^0(n+1,n)-(-\beta_L)\hat{z}_n-\left(-\beta_L(\hat{x}_n+D_L)-\gamma_0 D_L\right)\\ +(\gamma_b\hat{x}_n+b)-B_0\sin\omega_0 t_n\end{aligned} \quad \text{(B.11)}$$

$$\Omega_{R_2}(n+1,n)=A^0(n+1,n)-(\gamma_0\hat{z}_n)-(\gamma_0\hat{x}_n)+(\gamma_b\hat{x}_n+b)-B_0\sin\omega_0 t_n \quad \text{(B.12)}$$

$$\begin{aligned}\Omega_{R_3}(n+1,n) &= A^0(n+1,n)-(-\beta_R)\hat{z}_n-\left(-\beta_R(\hat{x}_n-D_R)+\gamma_0 D_R\right)\\ &\quad+(\gamma_b\hat{x}_n+b)-B_0\sin\omega_0 t_n\end{aligned} \tag{B.13}$$

$$A^n_{R_i}(n)=\begin{bmatrix}\hat{z}_n+\hat{x}_n+D_L & 0 & 0\\ D_L & -\hat{z}_n-\hat{x}_n & -D_R\\ 0 & 0 & \hat{z}_n+\hat{x}_n-D_R\\ \hat{x}_n & \hat{x}_n & \hat{x}_n\\ 1 & 1 & 1\\ -\sin\omega_0 t_n & -\sin\omega_0 t_n & -\sin\omega_0 t_n\end{bmatrix} \tag{B.14}$$

Unlike the approach developed in Ref. [19], in which each regime can be estimated independently, here the three regimes are coupled by the fact that the *y* nullcline is the same in all three regimes. This is analytically evident from the form of last three rows of the data matrix in Eq. (B.14), which couple all the columns together. Since loglikelihood maximization becomes a constrained quadratic problem, using vector formalism can help to adapt the notation to a form more suitable to quadratic constrained optimization. Consider the row vectors $\mathbf{A}_{n,i}=(A^1_{R_i}(n),\ldots,A^6_{R_i}(n))$ and $\boldsymbol{\psi}=(\psi_1,\ldots,\psi_6)$. If $\cdot$ denotes the scalar product, $\otimes$ the outer product, $'$ the transpose and $A_{n0}=A^0(n+1,n)$, each squared Ω_{R_i} can be written as

$$\begin{aligned}\left(\Omega_{R_i}(n+1,n)\right)^2 &= A^2_{n,0}+\left(\mathbf{A}_{n,i}\cdot\boldsymbol{\psi}'\right)^2+2A_{n,0}\left(\mathbf{A}_{n,i}\cdot\boldsymbol{\psi}'\right)\\ &= A^2_{n,0}+\frac{1}{2}\boldsymbol{\psi}K^1_{n,i}\boldsymbol{\psi}'+K^2_{n,i}\cdot\boldsymbol{\psi}'\end{aligned} \tag{B.15}$$

where

$$K^1_{n,i}=2\mathbf{A}'_{n,i}\otimes\mathbf{A}_{n,i},\qquad \mathbf{K}^2_{n,i}=2A_{n,0}\mathbf{A}_{n,i} \tag{B.16}$$

The McKean constraints from Eq. (18) are then written in the matrix form

$$\mathbf{K}^3\boldsymbol{\psi}'\le 0 \tag{B.17}$$

where

$$\mathbf{K}^3=\begin{bmatrix}0 & -D_L & 0 & D_L & -1 & 1\\ 0 & 1 & 0 & -1 & 0 & 0\\ -1 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & -1 & 0 & 0 & 0\\ 0 & -1 & 0 & 0 & 0 & 0\end{bmatrix} \tag{B.18}$$

Equations (B.15) and (B.17) are in a standard and very compact form that can be directly used in quadratic constrained numerical optimization solvers. If constraints are not necessary, an analytical solution to the loglikelihood maximization is easy to derive and computationally much faster to use. Unconstrained maximization of Eq. (B.2) in the 6 parameters implies 6 necessary conditions

$$\sum_{n=2}^{N} \Omega_R(n+1, n) \frac{\partial}{\partial \psi_j} \Omega_R(n+1, n) = 0, \qquad j = 1, ..., 6 \tag{B.19}$$

Equation (B.19) represents one 6 × 6 inhomogeneous algebraic system of six equations in the six unknown parameters, where the coefficients of each equation are expressed in terms of the elements of $A_R(n)$ and implicitly depend on the selection of the thresholds, Θ. For example, assume that $N = 2$ and that, for a given choice of thresholds, $\hat{x}_2 \in R_3$. Then, in Eq. (B.2) the loglikelihood sum has only one term, the term for $n = 2$. Recalling Eq. (B.10), the first, $j = 1$, condition becomes

$$\Omega_R \frac{\partial \Omega_R}{\partial \psi_1} = \Omega_{R_3} \frac{\partial \Omega_{R_3}}{\partial \psi_1} = \sum_{j=1}^{6} \left(A_{R_3}^{j}(n) A_{R_3}^{1}(n) \psi_j + \left(A_0(n+1, n) A_{R_3}^{1}(n) \right) \right) = 0 \tag{B.20}$$

because the terms $\Omega_{R_i} \dfrac{\partial \Omega_{R_i}}{\partial \psi_1}$ with $i \neq 3$ disappear due to the indicator functions. Since $A_{R_i}^{j}(n)$ and $A_0(n+1, n)$ are numerical data, Eq. (B.20) defines the first row

$$\sum_{j=1}^{6} C_j^1 \psi_j = -V^1 \tag{B.21}$$

of a nonhomogeneous matrix equation (i.e. a linear problem) in the ψ_j where the choice of thresholds defines the numerical value of C_j^1 and V^1. Taking into consideration a generic N and all js, in shorthand notation Eq. (B.21) becomes the first entry of the matrix equation

$$\mathbf{C}\boldsymbol{\psi}' = -\boldsymbol{\varepsilon}\mathbf{V}' \tag{B.22}$$

where **C** and the row vector **V** have components

$$C_k^l = \sum_{n=2}^{N} C_R(l, k; n) = \sum_{n=2}^{N} A_R^k(n) A_R^l(n) \tag{B.23a}$$

$$V^l = \sum_{n=2}^{N} h(n+1, n; \Delta t) A_R^l(n) \tag{B.23b}$$

where l, k = 1, ..., 6. Thus the estimate $\hat{\Psi}$ is obtained as

$$\boldsymbol{\psi}' = -\boldsymbol{\varepsilon}\mathbf{C}^{-1}\mathbf{V}' \tag{B.24}$$

Computationally, the coefficients C_k^j and V^l are obtained in the following way. For each C_k^j, at each n, find out which regime $\hat{x}_n$ belongs to, look up the appropriate column of the data matrix and substitute in the data. Then, sum up to N to find C_k^j. Do the same for V^j. Once the matrix **C** and the vector **V** have been computed, the linear system given by Eq. (B.22) can be inverted and solved to give the estimates $\hat{\beta}_L, \hat{\gamma}_0, \hat{\beta}_R, \hat{\gamma}_b, \hat{b}, \hat{B}_0$. Since these estimates are given by an inversion, scarcity of data for one regime affects the estimate for all the other regimes – a data set with many spikes generates a better estimate of the baseline regime than a data set with few spikes. Moreover, filtering the data for deseasonalization or spiking worsens these estimates. Notice that Eq. (B.22) is nonhomogeneous just because the parameter ε (which sets the time scale) is nonzero. If ε is considered to be a system parameter, uniqueness of solution is possible only when one of the other parameters is kept fixed (thus setting another scale). In the case $f = B_0 \sin \omega_0 t_n$, when $\Delta t = (2\pi)/(24\omega_0)$ there are 24 times t_n per period T_0. A multiple l_ϕ of a phase shift

$$\Delta\phi = \frac{\omega_0}{24} \tag{B.25}$$

can thus be added to the argument of the sine function as

$$f = B_0 \sin\left(\omega_0 t_n + l_\phi \Delta\phi\right) \tag{B.26}$$

where l_ϕ = 0, ..., 23. To avoid an additional search for the optimizing $\hat{l}_\phi$, the easiest thing to do is to exploit the fact that the AESO database reports demand data as well as price data. Since $f(t)$ models demand $d(t)$, Eq. (B.26) can be preliminarily and numerically synchronized with $d(t)$, by choosing the best l_ϕ. This can be carried out using power spectrum analysis or routines like finddelay in Matlab. In the case $f = 0$, it is more efficient to work directly with a smaller 5 × 3 data matrix, obtained from the matrix in Eq. (B.14) by removing the last row.

Loglikelihood maximization with respect to the variance σ^2 gives

$$\sigma^2 = \frac{\Delta t}{N-1}\sum_{n=2}^{N}\left(\Omega_R(n+1, n)\right)^2 \tag{B.27}$$

When the Ω_Rs in the sum are re-computed using the estimates of the parameters, they become the estimates $\hat{\Omega}_R$, and Eq. (B.27) gives the estimate $\hat{\sigma}^2$ of the variance σ^2.

Equation (B.24) shows that the estimates of the parameters are linear in ε. Looking at the form of $\Omega_{R_i}(n+1, n)$ in Eq. (B.8), it follows that $\hat{\Omega}_R(n+1, n)$ itself is linear in ε. Thus, Equation (B.27) implies that σ is linear in ε. Going back to Eq. (24b), it follows that ε cancels out in this equation. The same happens for Eq. (25). In other words, if the three mutually consistent assumptions are made that the noise is Gaussian and that the parameters and σ are linear in ε in the discrete time models of Eqs. (24) and Eq. (25) (and also in the continuous time models of Eqs. (20) and Eq. (9)), ε cancels out. This does not happen in the discretization of the original form of the McKean Eqs. (5) (nor in its discrete time form in Eqs. (A.1)). This consideration makes Eqs. (24) the best choice to model and calibrate electricity prices using a McKean mechanism.

As stated in Section 8, the estimation of the parameter σ has to be repeated on a 2-dimensional grid in the D_L, D_R plane. The $\hat{D}_L$, $\hat{D}_R$ pair which corresponds to the minimum σ^2 gives the best estimate [19] and maximizes the loglikelihood. There are obvious bounds on the values of D_L and D_R, such as $-D_L < D_R$. Moreover, $-D_L$ must be greater or equal to the minimum value of $\hat{x}_n$ found in the data set, and D_R must be less than or equal to the maximum $\hat{x}_n$. On a suitably shaped ($-D_L$, D_R) domain, a grid or a Montecarlo selection of pairs can then be explored, and this approach lends itself to code parallelization for faster calibration. One situation when this method cannot be applied is when some of the regimes do not capture data points. Imagine, for example, that a given D_L, D_R pair, regime 3 contains no data points. In this case, the third column of the data matrix $A_R(n)$ never appears in matrix C_k^j, for any n. The third elements $A_{R_1}^3(n)$ and $A_{R_2}^3(n)$ of the remaining second and third column of $A_R(n)$ are zero by definition, so that the third column C_3^j (j = 1, ..., 5) of C_k^j is zero, leaving C_k^j non--invertible. Neither β_R, nor any of the other parameters can be estimated. A way around this problem can be found by acknowledging the fact that in this case the underlying SETARX model consists of only two regimes and one threshold, so that the third row of the data matrix $A_R(n)$ and the third component of the vector V^j can be removed. This elimination produces a two regime problem that can now be estimated, the remaining two regimes still being coupled by the same y nullcline. The same happens when regime I contains no data points.

References

[1] AESO Procedures, 2012. Available on the AESO web site at URL http://www.aeso.ca/rulesprocedures/8778.html

[2] All market data used in this paper are available at the AESO web site, http://www.aeso.ca/

[3] ANDERSON C.L., DAVISON M., *A hybrid system-econometric model for electricity spot prices: considering spike sensitivity to forced outage distributions*, IEEE Trans. on Power Systems, 2008, 23, 3, 927–937.
[4] BARLOW M.T., *A diffusion model for electricity prices*, Mathematical Finance, 2002, 12, 4, 287–298.
[5] BECKER R., HUM A.S., PAVLOV V., *Modelling spikes in electricity prices*, Economic Record, 2007, 83, 263, 371–382.
[6] BENTH F.E., BENTH Y.S., KOEKEBAKKER S., *Stochastic modeling of electricity and related markets*, World Scientific, Singapore, 2008.
[7] BUNN D.W., OLIVERA F.S., *Agent-based simulation – An application to the new electricity trading arrangements of England and Wales*, IEEE Trans. on Evolutionary Computation, 2001, 5, 5, 493–503.
[8] CARTEA A., FIGUEROA M., *Pricing in electricity markets: A mean reverting jump diffusion model with seasonality*, A. Math. Fin., 2005, 12, 4, 313–335.
[9] CARTEA A., FIGUEROA M., GEMAN H., *Modelling electricity prices with forward looking capacity constraints*, App. Math. Fin., 2009, 16, 2, 103–122.
[10] COOMBES S., *Phase-locking in networks of pulse-coupled McKean relaxation oscillators*, Physica D, 2001, 160, 3–4, 173–188.
[11] DE JONG C., *The nature of power spikes: A regime-switch approach*, Studies in Nonlinear Dynamics and Econometrics, 2006, 10, 3.
[12] DE JONG C., HUISMAN R., *Option pricing for power prices with spikes*, Energy Power and Risk Management, 2003, 7, 11, 12–16.
[13] EYDELAND A., GEMAN H., *Fundamentals of energy derivatives*, [in:] *Energy modelling and the management of uncertainty*, Risk Books, London 1999, 35–43.
[14] GEMAN H., *Commodities and commodity derivatives*, Wiley, Chichester 2005.
[15] GEMAN H., RONCORONI A., *Understanding the fine structure of electricity prices*, Journal of Business, 2006, 79, 1225–1261.
[16] GŁADYSZ B., KUCHTA D., *Application of regression trees in the analysis of electricity load*, Operations Research and Decisions, 2008, 18, 4, 19–28.
[17] HAMILTON J.D., *A new approach to the economic analysis of nonstationary time series and the business cycle*, Econometrica, 1989, 57, 357–384.
[18] HAMILTON J.D., *Regime switching models*, [in:] S. N. Durlauf, L. E. Blume (Eds.), *The new Palgrave dictionary of economics*, 2nd Ed., Macmillan, Basingstoke 2008.
[19] HANSEN B.E., *Inference in TAR models*, Studies in Nonlinear Dynamics and Econometrics, 1997, 2, 1, 1–14.
[20] HUISMAN R., *The influence of temperature on spike probability in day-ahead power prices*, Energy Economics, 2008, 30, 5, 2697–2704.
[21] HUISMAN R., MATHIEU R., *Regime jumps in electricity prices*, Energy Economics, 2003, 25, 425–434.
[22] JANCZURA J., WERON R., *An empirical comparison of alternate regime-switching models for electricity spot prices*, Energy Economics, 2010, 32, 5, 1059–1073.
[23] JANJANI A., ZAREIPOUR H., *An overview of the operation of the Alberta electricity market*, Proceedings of the 23rd Canadian Conference on Electrical and Computer Engineering (CCECE), 2010, 1–6. Available on IEEE Xplore, DOI: 10.1109/CCECE.2010.5575104.
[24] KANAMURA T., OHASHI K., *On transition probabilities of regime switching in electricity prices*, Energy Economics, 2008, 30, 3, 1158–1172.
[25] KIM C.-J., NELSON C.R., *State space models with regime switching*, The MIT Press, Boston 1999.
[26] KUTOYANTS YU.A., *On identification of the threshold diffusion processes*, Annals of the Institute of Statistical Mathematics, 2012, 64, 2, 383–413.
[27] LUCHERONI C., *Resonating models for the electric power market*, Phys. Rev. E, 2007, 76, 056116/116.

[28] LUCHERONI C., *A resonating model for the power market and its calibration*, Proceedings of the 6th International Conference on the European Energy Market (EEM), 2009, 1–5. Available on IEEE Xplore, DOI: 10.1109/EEM.2009.5207157.
[29] LUCHERONI C., *Stochastic models of resonating markets*, Journal of Economic Interaction and Coordination, 2010, 5, 1, 77–88.
[30] LUCHERONI C., *TARX models for spikes and antispikes in electricity markets*, Proceedings of the 7th International Conference on the European Energy Market (EEM), 2010, 1–6. Available on IEEE Xplore, DOI: 10.1109/EEM.2010.5558724.
[31] MCKEAN H.P., *Nagumo's equation*, Adv. Math., 1970, 4, 209–223.
[32] MISIOREK A., TRUECK S., WERON R., *Point and interval forecasting of spot electricity prices: Linear vs. non-linear time series models*, Studies in Nonlinear Dynamics and Econometrics, 2006, 10, 3, 2.
[33] MOUNT T.D., NING Y., CAI K., *Predicting price spikes in electricity markets using a regime-switching model with time-varying parameters*, Energy Economics, 2006, 28, 1, 62–80.
[34] *Power Pool of Alberta*, Economic Withholding in the Alberta Energy Market, 2002.
[35] RAMBHARAT B.R., BROCKWELL A., SEPPI D., *A threshold autoregressive model for wholesale electricity prices*, J. R. Statist. Soc., 2005, Series C, 54, 287–299.
[36] SAPANKEVYCH N.I., SANKAR R., *Time series prediction using support vector machines: A survey*, IEEE Computational Intelligence Magazine, 2009, 4, 2, 24–38.
[37] See for example http://www.investopedia.com/terms/t/tight-market
[38] STOFT S., *Power system economics*, IEEE Press, Piscataway, 2003.
[39] TONG H., *Non-linear time series: A dynamical system approach*, Oxford University Press, Oxford 1990.
[40] TONG H., LIM K.S., *Treshold autoregression, limit cycles and cyclical data,* Journal of the Royal Statistical Society, 1980, Ser. B, 42, 245–292.
[41] TSAY R.S., *Testing and modeling threshold autoregressive processes*, J. Amer. Statist. Assoc., 1989, 84, 405, 231–240.
[42] TSAY R.S., *Testing and modelling multivariate threshold models*, J. Amer. Statist. Assoc., 1998, 93, 443, 1188–1202.
[43] *The Alberta power market*, W.D. Walls, A. Hollis (Eds.), The Van Horne Institute, Calgary, 2007.
[44] WERON R., *Modeling and forecasting electricity loads and prices: A statistical approach*, Wiley, Chichester, 2006.
[45] WERON R., MISIOREK A., *Forecasting spot electricity prices: A comparison of parametric and semi-parametric time series models*, International Journal of Forecasting, 2008, 24, 4, 744–763.
[46] WOOD A.J., WOLLENBERG B.F., *Power generation operation and control*, Wiley, Hoboken 1996.