Spikes, Antispikes and Thresholds in Electricity Logprices

Carlo Lucheroni
School of Science and Technologies,
University of Camerino,
via M. delle Carceri 9, 62032 Camerino (MC), Italy.
Email: carlo.lucheroni@unicam.it

***Index Terms*—Stochastic processes, nonlinear time series analysis, electricity market economics.**

EXTENDED ABSTRACT

I. INTRODUCTION

Electricity price data display features that are unusual in comparison to those of stock price data. Hourly electricity prices p_n show a complex mean reversion pattern, which includes upward spikes and a baseline periodicity. These features can be seen in Fig. 1, generated from data for the year 2006 from the Alberta (Canada) AESO energy market Ref.[1], where prices are shown in panel a). When data include exchanged volume, electricity prices display a very clear and strong correlation with volume levels (shown in Fig. 1 b)), whereas correlation of stock prices with volume is certainly not an obvious relation. Spikes appear only during daylight, always in coincidence with volume crests and never during night time. Even more interestingly, the mapping of prices into logprices reveals the presence of many antispikes (i.e. downward spikes), a feature that is not easily detectable directly from prices. As it can be seen in Fig. 1 c), antispikes appear only during night time, in coincidence with demand troughs. Antispikes have a heavy statistical weight, as shown in Fig. 2. Fig. 2 a) shows one year of logprices. Logprices mean revert around a base level, and jump very frequently upward and downward, with a quasi-symmetric, and somewhat skewed, aggregate distribution (see Fig. 2 b)) which has a bell shaped central part and thick tails. Since the Alberta market has a cap price and doesn't admit negative prices, both tails are limited in extension. Antispikes are not an artifact of the logaritmic mapping - which is just a way to emphasize them -, and due to their heavy statistical weight, they must be included in any econometric model that wants to model this kind of prices.

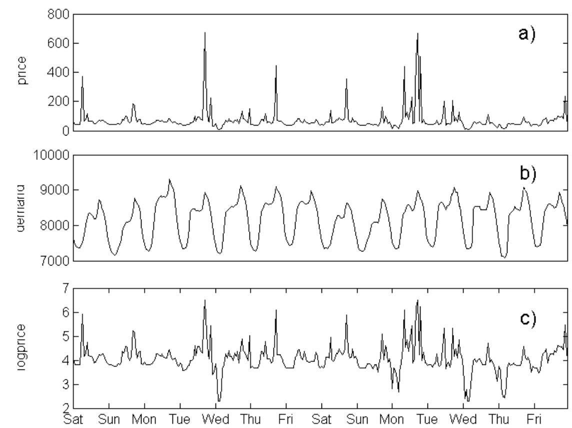

Fig. 1. Alberta power market prices: two weeks from Sat Nov-04-2006 to Sat Nov-17-2006, time in hours, time ticks in weekdays; a) prices in C$, b) demand in MWh, c) logprices.

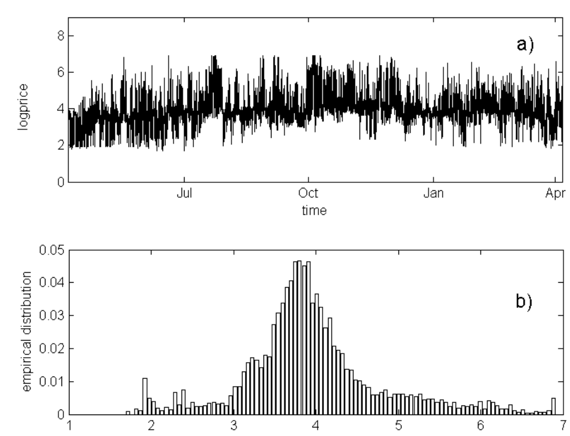

Fig. 2. Alberta power market prices: one year from Apr-07-2006 to Apr-06-2007; a) logprices, time in hours, time ticks in months b) (normalized) empirical distribution of logprices.

II. MODELS

Advanced econometric models for electricity data make use of continuous-time Lévy processes or discrete-time switching Markov chains to reproduce baseline and spikes, and almost never include antispikes. An alternative approach to model electricity data was sketched in Ref.[2], where it was shown that, both in continuous- and in discrete-time settings, the sinusoidal price baseline, the spikes and the antispikes can be modelled at the same time using a couple of sinusoidally driven piecewise linear stochastic differential or difference equations, i.e. using a Self-Exciting Threshold Vector Auto-Regression (SETVARX). In Ref.[2] this threshold model was discussed focussing on its remarkable dynamic property of being able, in a special parametric range, to sustain occasional firing of spikes in coincidence of demand crests and antispikes in coincidence of demand troughs, being demand modelled by a simple sinusoidal forcing term. In this model, it is the very interaction of nonlinearity with forcing and noise the combination that generates most interesting dynamic behavior. This SETVARX model can also be used in a more standard parametric range as a specially constrained TARX, useful to model series in which, besides spikes, threshold effects are considered important. In Ref.[3] a calibration technique appropriate for this model was developed, valid for every parametric regime. Another feature of the model is the fact that data must not be pre-filtered for spike identification, but simply fed directly to the calibration procedure. Since the Alberta data seem to contain strong threshold effects, as it will be now discussed, this SETVARX model can turn out to be very useful in studying the dynamic properties of the Alberta prices.

III. Data

It is a well known fact that electricity prices incorporate power stack effects. In the power industry, marginal costs increase nonlinearly with quantity, and the analytical relation between prices p_n (that in a perfect competition settings should be equal to marginal costs) and demand d_n (i.e. the exchanged volume) can be approximated by a 'golf stick' profile. This relationship can be catched by eye from inspection of the scatterplot shown in Fig. 3, where for each given time n the n-th data point (p_n, d_n) is plotted on the price/demand plane. On the left of the vertical line no price with level above that of the horizontal line exists, i.e. for a demand below that demand threshold, prices seem to be more or less following a linear price/demand relation. On the right of the demand threshold there is no clear price/demand relation, and for each demand value it seems that prices can either choose to follow demand linearly (remaining below the horizontal price threshold) or take arbitrarily high values, with some preference for the cap values. This means that above a certain demand, the prices can spike, but they don't necessarily do it, and when they spike they can take any value. Below that threshold demand, they never spike. In Fig. 4 for each given time n the n-th data point $(\log p_n, \Delta \log p_n)$ is plotted on the logprice / logprice-increment plane. There seem to exist at least three different regions in this plane, divided by the two vertical price threshold lines. Starting with a low logprice $\log p_n$ say on the left of the left hand side line, only moderate logprice increments $\Delta \log p_n$ can be attained. Starting with a high price on the right of the right line, higher increments can be attained. Starting between the two thresholds, low and high increments are allowed (the diagonal global shape depends on the existence of a cap and a floor for prices). Even more interestingly, when a Poincaré plot of successive price levels $(\log p_n, \log p_{n+1})$ is formed like that in Fig. 5, five regions divided by four thresholds become clearly visible. The graph is obviously symmetric, and it can be divided also in five horizontal regions, not shown in figure. It is apparent that when a logprice belonging to one of the vertical regions is recorded, the next logprice will be subject to strong and deterministic constraints in value. For example, if the starting logprice is located inside an antispike, i.e. in the second vertical region counting from left, the maximum increment it can take would move it to the third (central) horizontal region, i.e. near the baseline. This goes in accordance with the already mentioned fact that an antispike (which is fired only during night time) reverts always to the baseline, never transforming itself into a spike before daylight. The same in reverse would happen for spikes, but with larger flexibility (the fourth region is wider). All of this is also clear from direct inspection of the price series in time. What instead comes strikingly out from this Poincaré plot is the rigid thresholding of the dynamics. Logprices on the left and on the right of one threshold have a sharply different future. Restated in behavioral terms, investors seem to behave differently depending on the region of the price they observe. Any econometric model that wants to model this kind of prices must incorporate this structure.

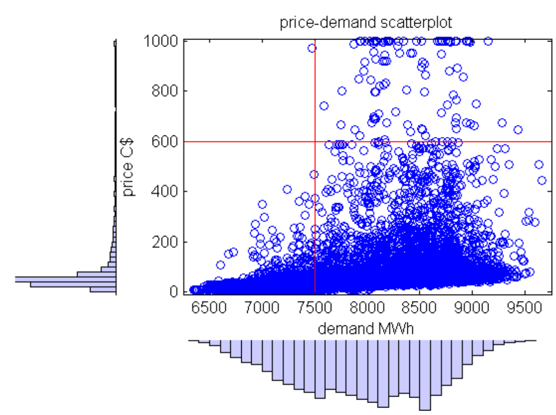

Fig. 3. 'Golf stick' power stack on the price/demand plane. Near each axis the marginal distribution of data is shown.

IV. Model

In the model of Ref.[2], in continuous time t, the logprice process $x(t) = \ln p(t)$ is the solution of the stochastic differential equations system

$$\epsilon \dot{x} = g_R(x) - y \quad (1a)$$
$$\dot{y} = x - \gamma_b\, y + b + f(t) + \sigma\, \xi(t), \quad (1b)$$

where $y(t)$ is a coordinate complementary to $x(t)$, $g_R(x)$ is a piecewise linear function of x, ϵ, γ_b, b and σ are constants, ξ is Normal noise and $f(t) = B_0 \sin \omega_0 t$. When the dynamics in

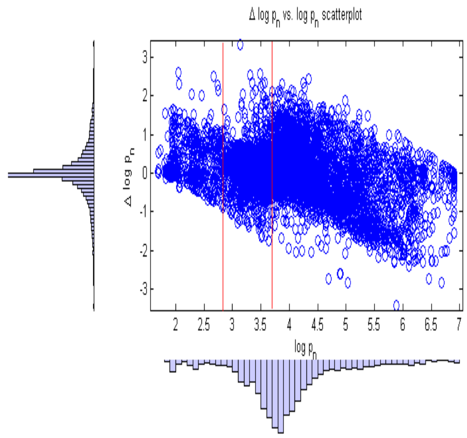

Fig. 4. Nonlinearity in logprice evolution: logprice increments.

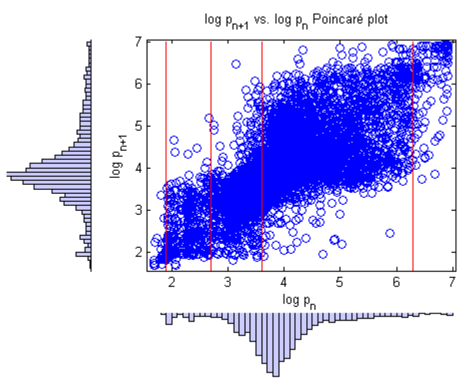

Fig. 5. Nonlinearity in logprice evolution: logprice constraints

Eqs. (1) is quiescent, i.e. $\sigma = f = 0$ and $\dot{x} = \dot{y} = 0$, Eqs. (1) read

$$y = g_R(x) \tag{2a}$$
$$y = (x + b)/\gamma_b. \tag{2b}$$

In the plane (x, y), i.e. in the phase-space of the dynamical system (1), Eqs. (2) can be drawn as two lines, called *nullclines*. In order to model spikes and antispikes $g_R(x)$ is chosen piecewise linear as

$$g_R(x) = \begin{cases} -\alpha_L(x + C_L) + \\ \gamma_0 D_L + \beta_L(-C_L + D_L), & -\infty < x \leq -C_L \\ & R = R_1 \\ \beta_L(x + D_L) + \gamma_0 D_L, & -C_L < x < -D_L \\ & R = R_2 \\ -\gamma_0 x, & -D_L \leq x \leq D_R \\ & R = R_3 \\ \beta_R(x - D_R) - \gamma_0 D_R, & D_R < x < C_R \\ & R = R_4 \\ -\alpha_R(x - C_R) - \\ \gamma_0 D_R + \beta_R(C_R - D_R), & C_R \leq x < +\infty \\ & R = R_5, \end{cases} \tag{3}$$

where α_L, α_R, β_L, β_R, γ_0, C_L, D_L, D_R, C_R are constants. The values of x where $g_R(x)$ forms kinks ($-C_L$, $-D_L$, D_R, C_R), are the model thresholds. This model contains four thresholds, as many as the scatterplot of Fig. 5 hints at. The five regions indicated by R_i, $i = 1, \ldots, 5$ are the five regimes of this threshold model. The special parametric range that sustains spikes and antispikes is for γ_b, α_L, α_R, β_L, β_R, γ_0 all greater than zero. In discrete time Eqs. (1) can be rewritten as

$$x_{n+1} - x_n = z_n \,\Delta t \tag{4a}$$
$$\begin{aligned} \epsilon\,(z_{n+1} - z_n) = {} & \left[\frac{\partial}{\partial x} g_R(x_n) - \epsilon\right] z_n \Delta t \\ & + [g_R(x_n) - (\gamma_b\, x_n + b) + f(t)]\,\Delta t \\ & - \sqrt{\Delta t}\,\sigma\,\eta_{n+1}, \end{aligned} \tag{4b}$$

where now $x_n = \log p_n$ is the discrete logprice at time step t_n, z_n is a complementary coordinate with the meaning of a logreturn intensity $\Delta \log p_n/\Delta t$, η_n is a normally distributed noise variable, $\omega_0 = 2\pi$ and Δt is set to $1/24$. This dynamics behaves in the same way as the dynamics of Eqs. (1), i.e. as a threshold vector autoregression where thresholds like $-D_L$ and D_R segment the system dynamics into sectors (regimes) and inside these sectors the dynamics is a standard (vector and linear) autoregression of lag 1 (very short and economic in terms of estimation time). A logprice x starting from one of the regimes R_i will follow a one-step-in-time evolution limited by the regime local dynamic rule. If x ends up in another regime, the next logprice will follow the next appropriate rule. Differently from what can be seen in Fig. 5, the model offers no limiting mechanism (capping or flooring) to the next values that x can take, unless it finds itself in the special parametric region. Anyway, in the generic parametric case, since within each regime the velocity with which x can escape from its original regime is regime-dependant, each regime has always its own characteristic range of possible outputs.

V. Calibration

When the SETVARX calibration procedure is fed with data, it will output four optimal threshold values and their associated estimated parameters. But since the data themselves seem to offer the proper threshold values, which are the positions of x that can be seen in Fig. 5, i.e. $-CL = 1.9$, $-DL = 2.7$, $DR = 3.6$, $CR = 6.3$, in this case the calibration will be run directly with these thresholds in order to see what happens (ϵ is set equal to $1/2$). In Fig. 6 the autocorrelation of the logprices is shown in panel a).

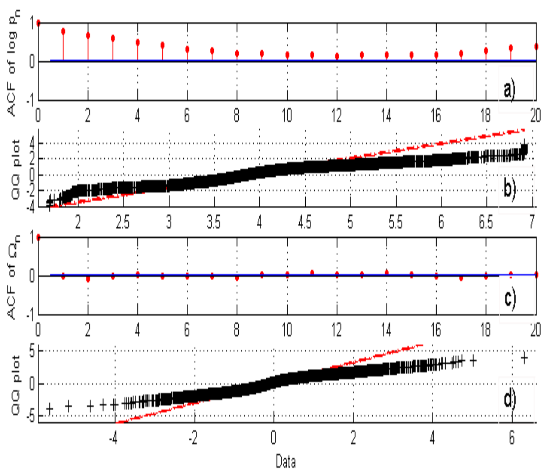

Fig. 6. ACF of logprices, QQ plot of their distribution, ACF of residuals, QQ plot of their distribution

The seasonality of demand results in seasonality of logprices. The QQ plot in panel b) puts in evidence the fact that the logprices have a non-normal distribution (see again Fig. 2 b)) - if they had it, the points would align on the dashed diagonal. The parameter values that are obtained from estimation are $\alpha_L = 14.5828$, $\beta_L = -14.1149$, $\gamma_0 = 15.0727$, $\beta_R = -16.6730$, $\alpha_R = 12.5268$, $\gamma_b = 28.4698$, $b = -171.8155$, $B_0 = -43.3043$. The signs of the estimates for β_L and β_R show that the model is outside the special parametric range, so that it behaves as a quasi-linear TARX. More surprisingly, the residuals r are practically uncorrelated, i.e. this simple five-regimes second order TARX model seems to capture most of the dynamic linear dependence in the data, even though not specifically designed for this. This can be seen in Fig. 6 c). Fig. 6 d) shows that the distribution of the residuals in strongly not normal. The shape of the residuals distribution is shown in Fig. 7 a) and their dynamics in Fig. 7 b). Another surprise comes from the fact that the residuals have a rather precise Double Gamma distribution. This can be seen in Fig. 8 a), where the right portion of the empirical distribution is fit by the distribution

$$p(r|u,s) = \frac{1}{u^s \Gamma(s)} r^{s-1} e^{-\frac{r}{u}}.$$

with $u = 139.0928$ and $s = 0.85619$. A fit by a purely exponential distribution, as displayed in Fig. 8 b), is not able to capture the kurtosis as well as the pole in the Gamma can do.

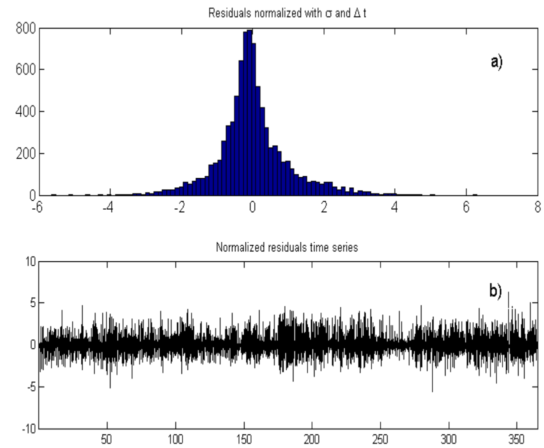

Fig. 7. Residuals from the estimation

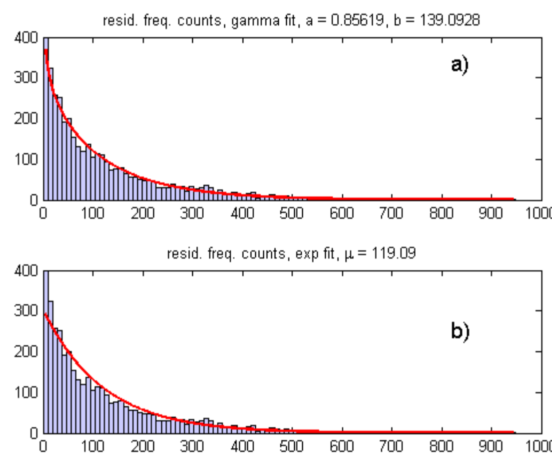

Fig. 8. Residuals distribution fits. a) Gamma, b) Exponential

Most of the model errors are then small errors. These errors are uncorrelated, but could have a higher order dynamic dependance. even though volatility clustering doesn't seem to be present in the series of Fig. 7 b). In facts, Fig. 9 a) shows that this is not much the case. In the Figure, the ACF of the squared residuals is computed. Some correlation indeed exists but it is very low for the first few lags (more or less 7 hours), and it is very small (even though not zero) at multiples of 24 hours. Fig. 9 b) shows the ACF of the squared logprices, to be compared with the ACF of the logprices. The two seem to be very similar, so that also the ACF of residuals and squared residuals can be expected very similar. An even weakly correlated squared residuals series suggests the possibility of GARCH effects, but a GARCH analysis of the residuals was carried out without any interesting outcome

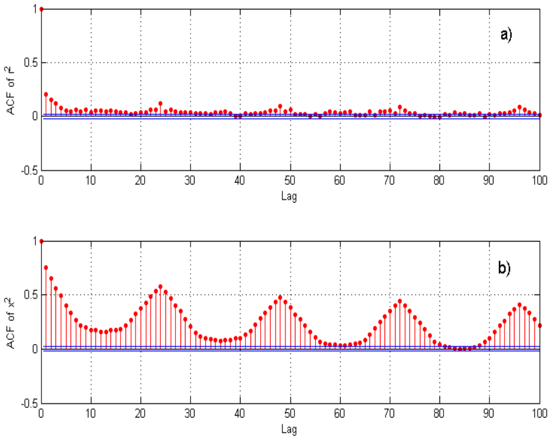

Fig. 9. ACF of a) squared residuals and b) squared loprices

- no GARCH effects are present. Some AR(q)X models with an exogenous sine forcing term were also tested on data, both on prices and logprices, up to $q = 6$, but it seems that the best model in terms of residuals uncorrelation is the SETVARX model itself, since ARX models are not able to remove correlation at all.

VI. Conclusions

To summarize, the main results of this analysis of the Alberta electricity data by means of the threshold model discussed in Ref.[2] seem to be the following. Besides spikes, antispikes are present in the data, and, what seems mostly important, there are four thresholds hidden in the data. After calibration of the model on data, it turns out that spikes are not generated by the mechanism that is at work in the special parametric regime of the threshold model. This notwithstanding, the threshold model in its generic parametric range can be anyway very effective in modelling such data in terms of removal of linear (and maybe higher) correlation from the residual series that it generates.

References

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

[2] C. Lucheroni, *TARX models for spikes and antispikes in electricity markets*, Proceedings of EEM 2010, IEEE Xplore DOI 10.1109/EEM.2010.5558724,

[3] C. Lucheroni, *Calibration of a SAS (Spike-AntiSpike) SETARX Model for Electricity Prices*, Proceedings of EEM 2012, in press.