On modelling of electricity spot price

Anna Nazarova, Rüdiger Kiesel and Fred Espen Benth

Institute of Energy Trading and Financial Services
University of Duisburg-Essen

Centre of Mathematics for Applications,
University of Oslo

25. August 2010
Models Setting
  Introduction
  Jump-diffusion model
  Threshold model
  Factor model

Implementation and Calibration
  Seasonality trend parameters estimation
  Jump-diffusion model calibration
  Threshold model calibration
  Factor model calibration

Assessing the models
  Models performance

Conclusions
Spot prices demonstrate such typical features as:

- *seasonality*: daily, weekly, monthly;
- *mean-reversion* or *stationarity*;
- *spikes*: may occur with some seasonal intensity;
- *high volatility*.

For our empirical analysis we use a data set of the Phelix Base electricity price index at the European Power Exchange (EEX).
EEX electricity spot price dynamics
Basics of the models

1. **Jump-diffusion model**, [1]. It proposes a one-factor mean-reversion jump-diffusion model, adjusted to incorporate seasonality effects.

2. **Threshold model**, [9] and [6]. It represents an exponential Ornstein-Uhlenbeck process driven by a Brownian motion and a state-dependent compound Poisson process.

3. **Factor model**, [3]. It is an additive linear model, where the price dynamics is a superposition of Ornstein-Uhlenbeck processes driven by subordinators to ensure positivity of the prices.
Basics of the models

Let \((\Omega, F, F_t, \mathbb{P})\) be a filtered probability space:

- time horizon \(t = 0, \ldots, T\) is fixed;
- in general, electricity spot price at time \(0 \leq t \leq T\) by \(S(t)\) takes the form:
  \[
  S(t) = e^{\mu(t)}X(t);
  \]
- \(\mu(t)\) is a deterministic function modelling the seasonal trend;
- \(X(t)\) is some stochastic process modelling the random fluctuation.
Basics of the models

Spot prices may vary with seasons:

$$\mu(t) = \alpha + \beta t + \gamma \cos(\epsilon + 2\pi t) + \delta \cos(\zeta + 4\pi t),$$ (2)

where the parameters $\alpha, \beta, \gamma, \delta, \epsilon$ and $\zeta$ are all constants:

- $\alpha$ is fixed cost linked to the power production;
- $\beta$ drives the long run linear trend in the total production cost;
- $\gamma, \delta, \epsilon$ and $\zeta$ construct periodicity by adding two maxima per year with possibly different magnitude.
Specification $X(t)$ for the jump-diffusion model

$$S(t) = e^{\mu(t)}X(t),$$
$$d \ln X(t) = -\alpha \ln X(t) \, dt + \sigma(t) \, dW(t) + \ln J \, dq(t),$$

- $\alpha$ is one mean-reversion parameter;
- $\sigma(t)$ is the time-dependent volatility;
- $J$ is the proportional random jump size, $\ln J \sim N(\mu_j, \sigma_j^2)$;
- $dq(t)$ is a Poisson process such that:

$$dq(t) = \begin{cases} 
1, \text{ with probability } ldt \\
0, \text{ with probability } 1 - ldt, 
\end{cases}$$

- $l$ is the intensity or frequency of spikes.
Specification $X(t)$ for the threshold model

\[ S(t) = e^{\mu(t)}X(t), \]
\[ d \ln X(t) = -\theta_1 \ln X(t) \, dt + \sigma \, dW(t) + h(\ln(X(t-))) \, dJ(t), \quad (3) \]

- $\theta_1$ is one mean-reversion parameter, positive constant;
- $\sigma$ is Brownian volatility parameter, positive constant.

The Brownian component models the normal random variations of the electricity price around its mean, i.e., *the base signal*. 
Specification $X(t)$ for the threshold model

\begin{align*}
S(t) &= e^{\mu(t)} X(t), \\
\ln X(t) &= -\theta_1 \ln X(t) dt + \sigma dW(t) + h(\ln(X(t^-))) dJ(t),
\end{align*}

where $J$ is a time-inhomogeneous compound Poisson process:

\[ J(t) = \sum_{i=1}^{N(t)} J_i, \]

and $N(t)$ counts the spikes up to time $t$ and is a Poisson process with time-dependent jump intensity. $J_1, J_2, \ldots$ model the magnitude of the spikes and are assumed to be i.i.d. random variables. The function $h$ attains two values, $\pm 1$, indicating the direction of the jump.
Specification $X(t)$ for the factor model

$$S(t) = e^{\mu(t)} X(t),$$

where $X(t)$ is a stochastic process represented as a weighted sum of $n$ independent non-Gaussian Ornstein-Uhlenbeck processes $Y_i(t)$

$$X(t) = \sum_{i=1}^{n} w_i Y_i(t),$$

where each $Y_i(t)$ is defined as

$$dY_i(t) = -\lambda_i Y_i(t) dt + dL_i(t), \quad Y_i(0) = y_i, i = 1, \ldots, n.$$  \hfill (5)

$w_i$ are weighted functions; $\lambda_i$ are mean-reversion coefficients; $L_i(t), t = 1, \ldots, n$ are independent càdlàg pure-jump additive processes with increasing paths.
Seasonality function $\mu(t)$ is common for both the factor and the threshold model. The method of *non-linear least squares* (OLS) is used for estimating the parameters.
Estimation parameters

- seasonality function, mean-reversion $\alpha$
- filtering spikes characteristics: size distribution and frequency
- rolling historical volatility $\sigma(t)$
Estimation parameters

- mean-reversion speed $\alpha$ is estimated by using linear regression, i.e. in the discrete version representation:

$$x_t = \theta_t + \beta x_{t-1} + \eta_t,$$

- $\theta_t$ is the mean-reverting level;
- $\beta$ is the modified mean-reversion speed;
- $\eta$ is the Brownian motion and jumps.
Estimation parameters

- extraction of the spikes from the original series of returns by simple iterative procedure;
- procedure repeats as long as no more outliers are filtered out;
- it gives the standard deviation of the jumps $\sigma_j$ and the cumulative frequency of jumps $I$;
Estimation parameters

rolling historical volatility is taken from Eydeland and Wolyniec [5]. \( m = 30 \) days, i.e. the width of the window:

\[
\sigma(t_k) = \sqrt{\frac{1}{m-1} \sum_{i=k-m+1}^{k} \left( \frac{\log P_i - \log P_{i-1}}{\sqrt{t_i - t_{i-1}}} - \sum_{i=k-m+1}^{k} \frac{\log P_i - \log P_{i-1}}{\sqrt{t_i - t_{i-1}}} \right)^2 } \tag{7}
\]
Estimation parameters
Estimation of the model parameters

- Jump threshold $\Gamma$ is set to filter out the jump and continuous paths.
- Then we estimate:
  - $\theta_1$ - the smooth mean-reversion force;
  - $\theta_2$ - the maximal expected number of jumps;
  - $\theta_3$ - the reciprocal average jump size;
  - $\sigma$ - the Brownian local volatility.
The approximative logarithmic likelihood function is constructed:

\[
L(\Theta | \Theta^0, E) = \sum_{i=0}^{n-1} \frac{(\mu(t_i) - E_i)\theta_1}{\sigma^2} \Delta E_i^c - \frac{\Delta t}{2} \sum_{i=0}^{n-1} \left( \frac{(\mu - E_i)\theta_1}{\sigma} \right)^2 \\
-(\theta_2 - 1) \sum_{i=0}^{n-1} s(t_i)\Delta t + N(t) \ln \theta_2 \\
+ \sum_{i=0}^{n-1} \left[ - (\theta_3 - 1) \frac{\Delta E_i^d}{h(E_i)} \right] + N(t) \ln \left( \frac{1 - e^{-\theta_3 \psi}}{\theta_3 (1 - e^{-\psi})} \right), \quad (8)
\]

it is possible to split it up into three independent parts and maximize them separately:

\[
L(\Theta | \Theta^0, E) = F_1(\theta_1) + F_2(\theta_2) + F_3(\theta_3). \quad (9)
\]
Procedure includes:

- deseasonalization
- identification the number of OU processes or factors involved
- filtering of the spike process and the base signal
- estimating of the base signal parameters
- analysis of the spike process
Assessment the number of factors required

Compare in the $L^2$ norm the empirical autocorrelation function (ACF) with theoretical one $\rho(k)$, see Barndorf-Nielsen and Shephard[2]:

$$
\rho(k) = \tilde{w}_1 e^{-k\lambda_1} + \tilde{w}_2 e^{-k\lambda_2} + \cdots + \tilde{w}_n e^{-\lambda n},
$$

(10)

where $k$ is a lag number; $\tilde{w}_i$ are positive constants summing up to 1; $\lambda_i$ are mean-reversion parameters. The larger $\lambda$, the faster the process comes back to its mean level, therefore it refers to the spike mean-reversion speed.

We obtained $n = 2$ as the optimal number of factors: one for spike and one for base signal.
Hard thresholding procedure

- is taken from Extreme Value Theory and helps to filter out the spikes;
- uses the methods from non-parametric statistics and provides as output both the base signal and the spike process;
- is reliable in the context of return distribution characteristics;

For details see *Meyer-Brandis and Tankov* [7] and *Nazarova* [8].
Base signal parameters estimation

- The main task here is to find the so-called *background driving* Lévy process $L_1(t)$ such that OU $Y_1(t)$ process has the same stationary distribution. A reasonable choice is the Gamma distribution, which is motivated by the opportunity to obtain an explicit analytical expression for the moments, otherwise compute them numerically, see *Barndorf-Nielsen and Shephard* [2].

- Evaluation of parameters of Gamma distribution is done by implementing a prediction-based estimating functions method developed by *Sørensen* [10] and *Bibby et al.* [4].
The method is closely related to the method of prediction error estimation that is used in the stochastic control literature. Here the method is applied to sums of Ornstein-Uhlenbeck processes. The estimating functions are based on predictors of functions of the observed process. We focus on a finite-dimensional space of predictors. For the optimal estimating functions this allows one to only involve unconditional moments.
Simulations comparison

EEX dynamics

Factor model

Threshold model

Jump-diffusion model
How do we assess the performance of the models?

- by simulating calibrated models dynamics;
- by computing the returns and their descriptive statistics;
- by comparing the moments of the returns with empirical ones.
### Method of moments

#### Tabelle: Comparative descriptive statistics results for the threshold, factor and jump-diffusion models.

<table>
<thead>
<tr>
<th>Moment</th>
<th>Average</th>
<th>Std. Dev</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>EEX</strong></td>
<td>0.0006</td>
<td>0.2985</td>
<td>0.4050</td>
<td>6.6179</td>
</tr>
<tr>
<td>Jump-diffusion model (Normal)</td>
<td>0.0007</td>
<td>0.3191</td>
<td>0.8343</td>
<td>10.3935</td>
</tr>
<tr>
<td>Threshold model (trunc. exp)</td>
<td>0.0006</td>
<td>0.2935</td>
<td>0.8336</td>
<td>5.9783</td>
</tr>
<tr>
<td>Factor model (Pareto)</td>
<td>0.0006</td>
<td>0.1595</td>
<td>1.6749</td>
<td>10.5308</td>
</tr>
<tr>
<td>Modified threshold model (Gamma)</td>
<td>0.0006</td>
<td>0.2822</td>
<td>0.5566</td>
<td>2.9946</td>
</tr>
<tr>
<td>Modified factor model (Gamma)</td>
<td>0.0006</td>
<td>0.1465</td>
<td>1.2414</td>
<td>5.7399</td>
</tr>
</tbody>
</table>
We have analysed and discussed the empirical performance of three continuous-time electricity spot price models.

Further investigation on the derivatives pricing.
Thank you for your attention!


