DEPARTMENT OF ECONOMICS
AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

Spikes and memory in (Nord Pool) electricity price spot prices

Tommaso Proietti, Niels Haldrup and Oskar Knapik

CREATES Research Paper 2017-39

Department of Economics and Business Economics
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

Spikes and memory in (Nord Pool) electricity price spot prices

Tommaso Proietti[*] University of Rome, 'Tor Vergata'

Niels Haldrup[†] Aarhus University

Oskar Knapik [‡] Aarhus University

November 13, 2017

Abstract

Electricity spot prices are subject to transitory sharp movements commonly referred to as spikes. The paper aims at assessing their effects on model based inferences and predictions, with reference to the Nord Pool power exchange. We identify a spike as a price value which deviates substantially from the *normal* price, where the latter is defined as the expectation arising from a model accounting for long memory at the zero and at the weekly seasonal frequencies, given the knowledge of the past realizations. Hence, a spike is associated to a time series innovation with size larger than a specified threshold. The latter regulates the robustness of the estimates of the underlying price level and it is chosen by a data driven procedure that focuses on the ability to predict future prices. The normal price is computed by a modified Kalman filter, which robustifies the inferences by cleaning the spikes, i.e. shrinking an observation deviating substantially from the normal price towards the one-step-ahead prediction. Our empirical application illustrates the effects of the spikes on the estimates of the parameters governing the persistence of the series; moreover, a real time rolling forecasting exercise is used to establish the amount of cleaning for optimizing the predicting accuracy at different horizons.

Keywords: Robustness; Kalman Filter; Long Memory.

JEL Classification: C22, C53, Q41.

[*]Corresponding author: Dipartimento di Economia e Finanza, via Columbia 2, 00133 Rome, Italy. E-mail: tommaso.proietti@uniroma2.it.

[†]Department of Economics and Business Economics, and CREATES - Center for Research in Econometric Analysis of Time Series, Fuglesangs Alle 4, 8210 Aarhus V, Denmark. E-mail: nhaldrup@econ.au.dk

[‡]Department of Economics and Business Economics, and CREATES - Center for Research in Econometric Analysis of Time Series, Fuglesangs Alle 4, 8210 Aarhus V, Denmark. E-mail: oknapik@econ.au.dk
The Authors acknowledge support from CREATES - Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation.

1 Introduction

Electricity spot prices are subject to transitory sharp movements commonly referred to as spikes. This phenomenon is well documented and defines one of the salient features of the corresponding time series. The literature has been looking for an operational definition of spikes and has formulated models that are capable of accounting for them. See for instance Hellström et al. (2012), Nowotarski et al. (2013), and the references therein. Another approach deals with robustifying the inferences, as in Dupuis (2017).

The objective of this paper is to assess the effects of spikes on model based inferences and predictions, with reference to spot electricity prices from the Nord Pool power exchange. In particular, we focus on the daily average price quoted in Elspot (Norwegian Krones), the auction market for day-ahead electricity delivery, see NordPool (2017). The series, which is available from January 1, 2000 to December 31, 2016 (6,210 daily observations), is plotted in the top left hand panel of figure 2, after a logarithmic transformation. We shall denote it by y_t, $t = 1, \ldots, n$. The top right panel displays the sample autocorrelations of y_t. Figure 1c plots the logarithmic returns $\Delta y_t = y_t - y_{t-1}$, and 1d is the corresponding sample autocorrelation function.

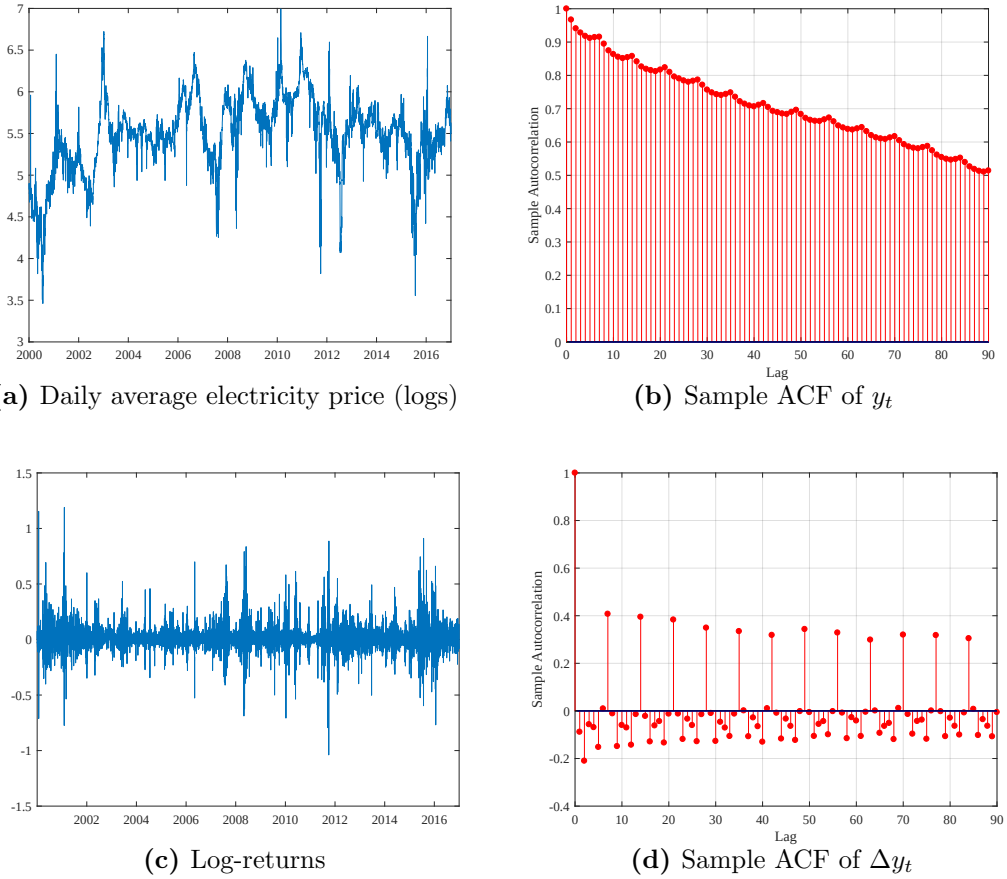
These plots reveal some of the most prominent features of the Nord Pool electricity prices: the strongly persistent movements of the underlying level, the presence of a weekly cycle, clearly visible in the autocorrelation plots. Interestingly, we could not detect an annual seasonal component. While few would question the occurrence of spikes, the literature has presented several operational definitions of a spike.

According to one popular definition, price spikes are extreme price values that surpass a specified absolute or relative threshold for a brief period of time. There are two relevant ingredients for characterizing a spike: the amplitude of the deviation from a reference value and the short duration. One possible limitation of this definition rests on the arbitrariness in defining the threshold value, which in turn can be periodically varying and functionally related to the price level itself, as well as to other covariates (temperature, etc) - in which case a conditional threshold would be preferable.

The approach that we take in this paper characterises a spike as a price value which deviates substantially from the *normal* price, where the latter is defined as the expectation arising from a correctly specified model of observed prices, given the knowledge of the past realizations. In particular, a spike is associated to a time series innovation with size larger than a specified threshold. The threshold, however, is not given, but rather it is backed up from the data by performing a real time forecasting exercise and establishing the level of robustification against spike that is optimal for forecasting any number of days ahead. As in Hellström et al. (2012) we allow for both positive and negative spikes.

The model that vehicles the notion of a normal price is a generalized fractional exponential model formulated in the frequency domain. According to it the logarithm of the spectral density is additively decomposed into the contribution of the long memory component at the zero frequency, that of a persistent weekly cycle and that of the short memory component. More details will be provided in section 2.

Under the Gaussian assumption, the model is estimated by maximising the Whittle likelihood, as it is described in section 3.1 and prediction can be carried out by using the autoregressive representation of the model. However, the presence of spikes makes the



(a) Daily average electricity price (logs)

(b) Sample ACF of y_t

(c) Log-returns

(d) Sample ACF of Δy_t

Figure 1: Nord Pool average daily prices (logs): original series and sample autocorrelation function; log-returns and their sample autocorrelation function.

assumption of Gaussianity untenable and will distort inferences and predictions.

Hence, the central contribution of our paper deals with proposing a methodology for robustifying the inferences through a modified Kalman filter, which identifies when the innovation at time t has a size greater than a threshold and limits the influence of the corresponding observation by shrinking it towards its one-step-ahead prediction, conditional on the previous observations. This idea is developed in section 4. The cleaned time series data can then be used for model based inferences and the prediction of future observations.

Our application to the Nord Pool electricity price series illustrates the effects of the spikes on the estimates of the parameters governing the persistence of the series. A real time rolling forecasting exercise enables to determine the level of correction that is optimal for predicting at different horizons. The main results will be presented in section 5. In section 6 we present some conclusions.

2 A Generalized Fractional Exponential Model for Electricity Prices

We consider the following time series regression model for the logarithmic daily average price y_t:

$$y_t = x_t'\beta + u_t, \tag{1}$$

where x_t is a $k \times 1$ vector of explanatory variables and u_t is a zero mean random process. Denoting by B the lag operator, u_t is generated by the following fractionally integrated process

$$(1-B)^{d_0} \prod_{j=1}^{3} (1 - 2\cos\varpi_j B + B^2)^{d_j} u_t = z_t, \quad \varpi_j = \frac{2\pi}{7} j, j = 1, 2, 3, \tag{2}$$

where $d_j, j = 1, 2, 3$ are the fractional integration parameters at the Gegenbauer frequencies and z_t is a short-memory stationary process, characterized by the spectral density $f_z(\omega)$, which will be defined below. The factor $(1-B)^{d_0}$ accounts for the long-range dependence at the long-run frequency, whereas the Gegenbauer polynomials $(1-2\cos\varpi_j B+B^2), j = 1, 2, 3$, account for the persistent seasonal behavior of the process at frequencies $\varpi_j = \frac{2\pi}{7} j$, which correspond to cycles with a period of 7 days ($j = 1$), 3.5 days ($j = 2$, two cycles per week), and 2.3 days ($j = 3$, three cycles per week). The process is stationary if $d_j \in (0, 0.5)$, $j = 0, 1, 2, 3$ (see Woodward et al. (2011), p. 418, Theorem 11.5a). The Gegenbauer process was introduced by Hosking (1981) and formalized by Gray et al. (1989). The model can be considered as an extension of that specified by Hsu and Tsai (2009).

As far as the short-memory component z_t is concerned, we assume that its spectral density has the following exponential representation (see Bloomfield (1973)) of order q:

$$f_z(\omega) = \frac{1}{2\pi} \exp\left(c_{z0} + 2\sum_{k=1}^{q} c_{zk} \cos(\omega k) \right),$$

where $c_{z0}, c_{zk}, k = 1, \ldots, q$ are real-valued parameters, known as the *cepstral* coefficients of z_t Bogert et al. (1963).

Our specification is similar to that adopted by Soares and Souza (2006), but it differs for the representation of the short memory components, which is modelled by truncating the Fourier series representation of the log-spectrum; furthermore, we do not constrain the memory parameters to lie in the stationary region.

The specification implies that the spectral density of u_t, denoted as $f(\omega)$, is linear in the memory coefficients and in the short-run coefficients c_{zk}, as can be seen by writing:

$$\begin{aligned} \ln[2\pi f(\omega)] \quad = \quad & c_{z0} + 2\textstyle\sum_{k=1}^{q} c_{zk}\cos(\omega k) - 2d_0 \ln|2\sin(\omega/2)| \\ & -2\textstyle\sum_{j=1}^{3} d_j \ln\left|4\sin\left(\frac{\omega+\varpi_j}{2}\right)\sin\left(\frac{\omega-\varpi_j}{2}\right)\right|. \end{aligned} \tag{3}$$

From (3) we can derive the cepstral coefficients of the process u_t, by the inverse Fourier transform:

$$c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln[2\pi f(\omega)] \exp(\imath\omega k) \mathrm{d}\omega, k = 0, 1, \ldots, \tag{4}$$

which, by trigonometric identities (see Gradshteyn and Ryzhik (2007), formula 1.441.2 and 1.448.2), are given by

$$c_k = I(k \leq q)c_{zk} + \frac{1}{k}\left(d_0 + 2\sum_{j=1}^{3} d_j \cos(\varpi_j k)\right), \quad k \geq 1, \tag{5}$$

whereas, for $k = 0$, $c_0 = c_{z0}$. The sequence $\{c_k\}_{k=0,1,\ldots}$, is referred to as the *cepstrum* of u_t. Its elements are the Fourier coefficients of the log-spectrum of u_t, i.e. $\ln[2\pi f(\omega)] = c_0 + 2\sum_{k=1}^{\infty} c_k \cos(\omega k)$, as it is clear from (4). The cepstral coefficient contain all of the relevant information that is needed for the linear prediction of the process u_t, and y_t thereof, as we shall see in section 3.2.

3 Estimation and prediction

3.1 Whittle estimation

Given a time series realization of length n, $\{(y_t, x_t), t = 1, 2, \ldots, n\}$, estimation of the memory parameters $d_j, j = 0, 1, 2, 3$, and the cepstral coefficients $c_{zk}, k = 0, 1, \ldots, q$, is carried out in the frequency domain.

Let $\hat{\beta}$ denote the least squares estimator of β and define $\hat{u}_t = y_t - x_t'\hat{\beta}$. The periodogram of $\hat{u}_t$ is

$$I(\omega) = \frac{1}{2\pi n}\left|\sum_{t=1}^{n} \hat{u}_t e^{-\imath\omega t}\right|^2.$$

Defining $\theta = (d_0, d_1, d_2, d_3, c_{z0}, c_{z1}, \ldots, c_{zq})'$, the Whittle likelihood is

$$\mathcal{L}(\theta) = -\sum_{j=1}^{n-1}\left[\ln f(\omega_j) + \frac{I(\omega_j)}{f(\omega_j)}\right]. \tag{6}$$

where $f(\omega)$, the spectral density of u_t, is specified according to equation (3),and both the periodogram and the spectrum are evaluated at the Fourier frequencies $\omega_j = \frac{2\pi j}{n}$, $j = 1, \ldots, n-1$.

The maximizer of (6) is the Whittle pseudo-maximum likelihood estimator of θ. We refer to Dahlhaus (1989), Velasco and Robinson (2000), Giraitis et al. (2012) and Beran et al. (2013) for the properties of the estimator in the long memory case.

For estimation we use the tapered periodogram. Tapering aims at reducing the estimation bias that characterizes the periodogram ordinates in the nonstationary case. Velasco (1995) and Velasco and Robinson (2000) show that with adequate data tapers, the Whittle estimator of the parameters of classes of fractional integrated models, encompassing the fractional exponential model, is consistent and asymptotically normal when the true memory parameter is in the nonstationary region. The adoption of a data taper makes the estimates invariant to the presence of certain deterministic trends.

The tapered discrete Fourier transform of y_t is defined as the squared modulus of

$$w(\omega_j) = \left(2\pi\sum_{t=1}^{n} h_t^2\right)^{-1/2} \sum_{t=1}^{n} h_t \hat{u}_t e^{i\omega_j t} \tag{7}$$

where $\{h_t\}_{t=1}^n$ is a taper sequence, i.e. a sequence of nonnegative weights that down-weight the extreme values of the sequence on both sides, leaving the central part almost unchanged. Note that the raw periodogram arises in the case of $h_t = 1$. As in Velasco and Robinson (2000), the sequence $\{h_t\}_{t=1}^n$ is obtained from the coefficients of the polynomial

$$\left(\frac{1 - z^{[n/p]}}{1 - z}\right)^p.$$

Typical choices are $p = 2, 3$. The tapered periodogram that enters the Whittle likelihood in (6) is then $I(\omega_j) = |w(\omega_j)|^2$.

3.2 Linear representation and prediction

Let $\mathcal{F}_t$ denote the information available up to time t. The one-step-ahead prediction error variance (p.e.v.) of y_t, $\sigma^2 = Var(y_t|\mathcal{F}_{t-1})$, is obtained by the Szegö-Kolmogorov formula as

$$\sigma^2 = \exp\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln[2\pi f(\omega)]\mathrm{d}\omega\right],$$

from which it follows that $c_{z0} = \ln \sigma^2$. Moreover, if the stationarity condition is satisfied, we can obtain the infinite moving average (MA) and autoregressive (AR) representations of the system:

$$y_t = x_t'\beta + \psi(B)\xi_t, \quad \phi(B)(y_t - x_t'\beta) = \xi_t, \quad \xi_t \sim WN(0, \sigma^2), \tag{8}$$

where $\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \ldots, \sum_j \psi_j^2 < \infty$, and $\phi(B) = \sum_{j=0}^{\infty} \phi_j B^j = \psi(B)^{-1}$, $\sum_j \phi_j^2 < \infty$, $\phi_0 = 1$. The moving average coefficients of the Wold representation are obtained recursively from the cepstral coefficients by the formula

$$\psi_j = j^{-1}\sum_{r=1}^{j} r c_r \psi_{j-r}, \qquad j = 1, 2, \ldots, \tag{9}$$

with $\psi_0 = 1$. See Jancek (1982), Pourahmadi (1983) and Hurvich (2002). The autoregressive coefficients are obtained according to the recursive formula

$$\phi_j = -j^{-1}\sum_{r=1}^{j} r c_r \phi_{j-r}, \qquad j = 1, 2, \ldots, \tag{10}$$

with starting value $\phi_0 = 1$. The optimal one-step-ahead linear predictor of the random component is then $\tilde{u}_{t+1|t} = \sum_{j=1}^{\infty} \phi_j u_{t-j+1}$.

In practice, the infinite MA and AR representations are truncated at lag m, and the approximating AR(m) or MA(m) models can be cast in state space form. In general, an ARMA(p, q) time series model for u_t,

$$u_t - \phi_1 u_{t-1} - \cdots - \phi_p u_{t-p} + \xi_t + \theta_1 \xi_{t-1} + \cdots + \theta_q \xi_{t-q}, \xi_t \sim \mathrm{WN}(0, \sigma^2),$$

can be written in state space form with measurement equation:

$$u_t = Z\alpha_t + G\xi_t, \qquad t = 1, \ldots, n, \tag{11}$$

where α_t is a random vector with $m = \max(p, q)$ elements, $Z = [1, 0, \ldots, 0]$, and $G = 1$. The evolution of the states is governed by the transition equation

$$\alpha_{t+1} = T\alpha_t + H\xi_t, \qquad t = 1, 2, \ldots, n, \tag{12}$$

where

$$T = \begin{bmatrix} \phi_1 & 1 & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \vdots & \cdots & \cdots & 0 & 1 \\ \phi_m & 0 & \cdots & \cdots & 0 \end{bmatrix}, H = \begin{bmatrix} \theta_1 + \phi_1 \\ \theta_2 + \phi_2 \\ \vdots \\ \vdots \\ \theta_m + \phi_m \end{bmatrix}.$$

The initial state vector α_1, assuming stationarity (the eigenvalues of T are inside the unit circle), has a distribution with mean $\mathrm{E}(\alpha_1) = 0$ and variance $\mathrm{Var}(\alpha_1) = \sigma^2 P_{1|0}$, satisfying matrix equation $P_{1|0} = TP_{1|0}T' + HH'$. The above state space representation is due to Pearlman (1980), and encompasses both the pure AR and MA cases. An approximation of the optimal predictor using m lags is then obtained from the Kalman filter.

Define $\tilde{\alpha}_{t|t-1} = \mathrm{E}(\alpha_t|\mathcal{F}_{t-1})$, and $\mathrm{Var}(\alpha_t|\mathcal{F}_{t-1}) = \sigma^2 P_{t|t-1}$. The Kalman filter (KF) is the following recursive algorithm: for $t = 1, \ldots, n$,

$$\begin{array}{ll} \nu_t = u_t - Z\tilde{\alpha}_{t|t-1}, & f_t = ZP_{t|t-1}Z' + GG', \\ & C_t = P_{t|t-1}Z' f_t^{-1}, \\ \tilde{\alpha}_{t|t} = \tilde{\alpha}_{t|t-1} + C_t\nu_t, & P_{t|t} = P_{t|t-1} - C_t f_t C_t', \\ & Q_t = HG' f_t^{-1}, \\ \tilde{\alpha}_{t+1|t} = T\tilde{\alpha}_{t|t} + Q_t\nu_t, & P_{t+1|t} = TP_{t|t}T' + HH' - (Q_t f_t Q_t' + Q_t f_t C_t' T' + TC_t f_t Q_t'). \end{array} \tag{13}$$

As $m \to \infty$, the above equations compute the innovations $\nu_t = u_t - \mathrm{E}(u_t|\mathcal{F}_{t-1})$, and $\sigma^2 f_t$ is the prediction error variance at time t, which is $\mathrm{Var}(u_t|\mathcal{F}_{t-1})$; $\tilde{\alpha}_{t|t}$ are the updated, or real-time, estimates of the state vector, and C_t is the gain, $C_t = \mathrm{Cov}(\alpha_t, y_t|Y_{t-1})[\mathrm{Var}(y_t|Y_{t-1})]^{-1}$. It can be shown that $\alpha_t|\mathcal{F}_t \sim \mathrm{N}(\tilde{\alpha}_{t|t}, \sigma^2 P_{t|t})$. The vector Q_t has the following interpretation:

$$Q_t = \mathrm{Cov}(H\xi_t, u_t|\mathcal{F}_{t-1})[\mathrm{Var}(u_t|\mathcal{F}_{t-1})]^{-1},$$

so that $\tilde{\alpha}_{t+1|t}$ is the one-step-ahead state prediction and we can write $\alpha_{t+1}|\mathcal{F}_t \sim \mathrm{N}(\tilde{\alpha}_{t+1|t}, \sigma^2 P_{t+1|t})$.

4 Robust filtering and forecasting

As stated in the Introduction, electricity prices are characterized by abnormally high or low values, called price spikes. Their effect on the periodogram and, thus, on the Whittle estimates depends on their size, pattern and recurrence. Spikes tend to bias downward the estimates of the memory parameters, and to inflate those of the conditional and unconditional variance of the series. They affect the predictions from the model, not only indirectly, via their effect on the parameter estimates, but also directly, as they have a potentially unbounded effect on the predictions.

There are several strategies for robustifying the estimates of the parameters and, thus, of the spectrum. McCloskey and Hill (2013) propose to replace the sample spectrum $I(\omega)$

in (6) with a robust periodogram, constructed from the Fourier transform of a robust autocovariance estimate. Our alternative strategy is based on an iterative data-cleaning method, based on a robust Kalman filter, exposed below. The robust KF is a data cleaning algorithm based on Martin and Thomson (1982), which reduces the influence of outliers (and the spikes) by replacing the outlying observations with a regularized estimate, resulting from shrinking y_t towards its one-step-ahead prediction, by an amount which depends on the size of the innovation.

We now propose a robust Kalman filter which is an adaptation of the data cleaning filter proposed by Masreliez and Martin (1977) and Martin and Thomson (1982). The robust filter modifies the updating and prediction equations (13), by using a bounded and continuous function of the standardized innovations, so as to control the effects of extreme values on the conditional mean estimates and the predictions.

Let $\Psi(x)$ denote Huber's influence function, see Maronna et al. (2006):

$$\Psi(x) = \begin{cases} x, & \text{if } |x| \leq a, \\ a \cdot \text{sign}(x), & \text{if } |x| > a. \end{cases}$$

This function limits the influence of a large x, by replacing it by a constant value a bearing the same sign. The corresponding weight function is $w(x) = \Psi(x)/x$, such that $0 \leq w(x) \leq 1$.

The robust KF is obtained by replacing in the usual KF f_t^{-1} by a shrunk version $\bar{f}_t^{-1}$ in the expressions for the regression matrices C_t and Q_t, giving, for $t = 1, \ldots, n$:

$$\begin{array}{ll} \nu_t^+ = u_t - Z\tilde{\alpha}_{t|t-1}^+, & f_t^+ = ZP_{t|t-1}^+ Z' + GG', \;\; \bar{f}_t^{-1} = w\left(\nu_t^+/\sqrt{f_t^+}\right)/f_t^+ \\ & C_t^+ = P_{t|t-1}^+ Z' \bar{f}_t^{-1}, \\ \tilde{\alpha}_{t|t}^+ = \tilde{\alpha}_{t|t-1}^+ + C_t^+ \nu_t^+, & P_{t|t}^+ = P_{t|t-1}^+ - C_t^+ \bar{f}_t C_t^{+\prime}, \\ & Q_t^+ = HG' \bar{f}_t^{-1}, \\ \tilde{\alpha}_{t+1|t}^+ = T\tilde{\alpha}_{t|t}^+ + Q_t^+ \nu_t^+, & P_{t+1|t}^+ = TP_{t|t}^+ T' + HH' - (Q_t^+ \bar{f}_t Q_t^{+\prime} + Q_t^+ \bar{f}_t C_t^{+\prime} T' + TC_t^+ \bar{f}_t Q_t^{+\prime}). \end{array} \tag{14}$$

Here, $\tilde{\alpha}_{t|t-1}^+$ and $P_{t+1|t}^+$ are initialized at time $t_0 \geq 1$ by $\tilde{\alpha}_{t_0|t_0-1}^+ = \tilde{\alpha}_{t_0|t_0-1}$ and $P_{t_0+1|t_0}^+ = P_{t_0+1|t_0}$. Note that if $\Psi(x) = x$, the identity function, $w(x) = 1$ and the above recursions yield the usual Kalman filter (13).

The Huber Ψ-function applies to the standardized innovation $\nu_t^+/\sqrt{f_t^+}$; if its absolute value is larger than a, then the weight function $w\left(\nu_t^+/\sqrt{f_t^+}\right)$ is less than 1 and $\bar{f}_t^{-1}$ is shrunk towards zero. The "clean" estimate of y_t is then $\tilde{y}_{t|t} = x_t'\beta + \tilde{u}_{t|t}$, where

$$\tilde{u}_{t|t} = Z\tilde{\alpha}_{t|t}^+ + GG' \bar{f}_t^{-1} \nu_t^+. \tag{15}$$

Assume that all the previous observations were clean, so that $w(\tilde{\nu}_s) = 1, s < t$. At time t, $\nu_t^+ = \nu_t, \alpha_{t|t-1}^+ = \alpha_{t|t-1}$, and $f_t^+ = f_t$, whereas $\tilde{\alpha}_{t|t}^+$ is shrunk towards the one-step-ahead prediction $\alpha_{t|t-1}$. If $w\left(\nu_t^+/\sqrt{f_t^+}\right) = 0$, then no updating takes place. Noticing that $u_t = Z\tilde{\alpha}_{t|t} + GG' f_t^{-1} \nu_t$, we can write

$$\tilde{u}_{t|t} = w\left(\nu_t/\sqrt{f}_t\right) u_t + \left[1 - w\left(\nu_t/\sqrt{f}_t\right)\right] Z\tilde{\alpha}_{t|t-1},$$

a weighted average of the observable u_t and its one-step-ahead prediction using the past observations. Similarly, the cleaned observation $\tilde{y}_{t|t}$ is intepreted as a weighted average of y_t and the one-step-ahead prediction, $x_t'\beta + Z\tilde{\alpha}_{t|t-1}$.

A spike is defined in real time as the deviation of y_t from its cleaned value $\tilde{y}_{t|t}$, i.e. $s_t = y_t - \tilde{y}_{t|t}$. Obviously, the sequence s_t depends on the choice of the tuning constant a. In robust statistics, see Maronna et al. (2006), the latter regulates the trade–off between the so–called breakdown point and the efficiency of the estimator. The breakdown point is a measure of robustness of an estimator as it gives the fraction of bad data the estimator can tolerate before giving results towards the boundary of the parameter space. Lower values of a increase the breakdown point, but reduce the efficiency. A typical value is $a = 1.345$ which guarantees 95% efficiency for i.i.d. Gaussian observations.

Robust estimation and forecasting entails the following steps:

1. The model (2)-(3) is estimated by maximising the Whittle likelihood in (6) with respect to θ, using a taper of order $p = 2$ for the periodograms. Let $\hat{\theta} = (\hat{\beta}, \hat{d}_0, \hat{d}_1, \hat{d}_2, \hat{d}_3, \hat{c}_{z0}, \ldots, \hat{c}_{zq})'$ denote the estimated coefficients.

2. We obtain from $\hat{\theta}$ the estimated cepstral coefficients $\hat{c}_k, k = 1, 2, \ldots, m$, applying (5), and we then obtain the first m coefficient of the autoregressive representation, $\hat{\phi}_j, j = 1, \ldots, K$, by applying formula (10).

3. The truncated AR(m) representation for u_t, $u_t = \sum_{j=1}^{m} \hat{\phi}_j u_{t-j} + \xi_t$, where ξ_t is the error term, is cast in state space form.

4. The robust Kalman filter (14) is applied in order to obtain a set of clean data.

After some experimentation, we set $m = 50$ for the order of the AR approximation. The selection of the order q of the EXP model for z_t is based on the Bayesian Information Criterion (BIC) and it is performed in the first step of the algorithm. If $\tilde{\phi}_j, j = 1, \ldots, K$, and $\tilde{\beta}$ denote the final estimates of the AR and the regression coefficients, our predictor of $y_{t+l}, l = 1, 2, \ldots,$ at time t is obtained as $\tilde{y}_{t+l|t} = x_{t+l}'\tilde{\beta} + \tilde{u}^{+}_{t+l|t}$, where $\tilde{u}^{+}_{t+l|t}$ is the l-step ahead predictor of u_t obtained from applying the robust Kalman filter to the state space representation of the AR(m) model.

The steps 1-4 can also be repeated, after replacing the observations y_t by the cleaned observations $\tilde{y}_{t|t}$ as given in equation (15), until the convergence is achieved, i.e. the change in the parameter estimates are within the prescribed tolerance.

5 Case study: Nord Pool daily electricity prices

The model was estimated using the complete data available for the period 1/1/2000-31/12/2016, for a total of 6,210 obervations. Table 1 displays the parameter estimates, along with their standard errors, arising from Whittle maximum likelihood estimation using the original uncorrected data, the clean data, applying the robust Kalman filter using the Huber influence function with tuning constant set equal to $a = 1.345$, with is standard reference value. In the last two columns we present the estimates arising from iteratively replacing the cleaned data in Whittle estimation. The procedure converged

in 6 iteration; convergence is achieved when from one iteration to the next the sum of the squared differences in the weights $w(\cdot)$ is less that 10^{-4}. The order of the short run component estimated by BIC is $q = 7$.

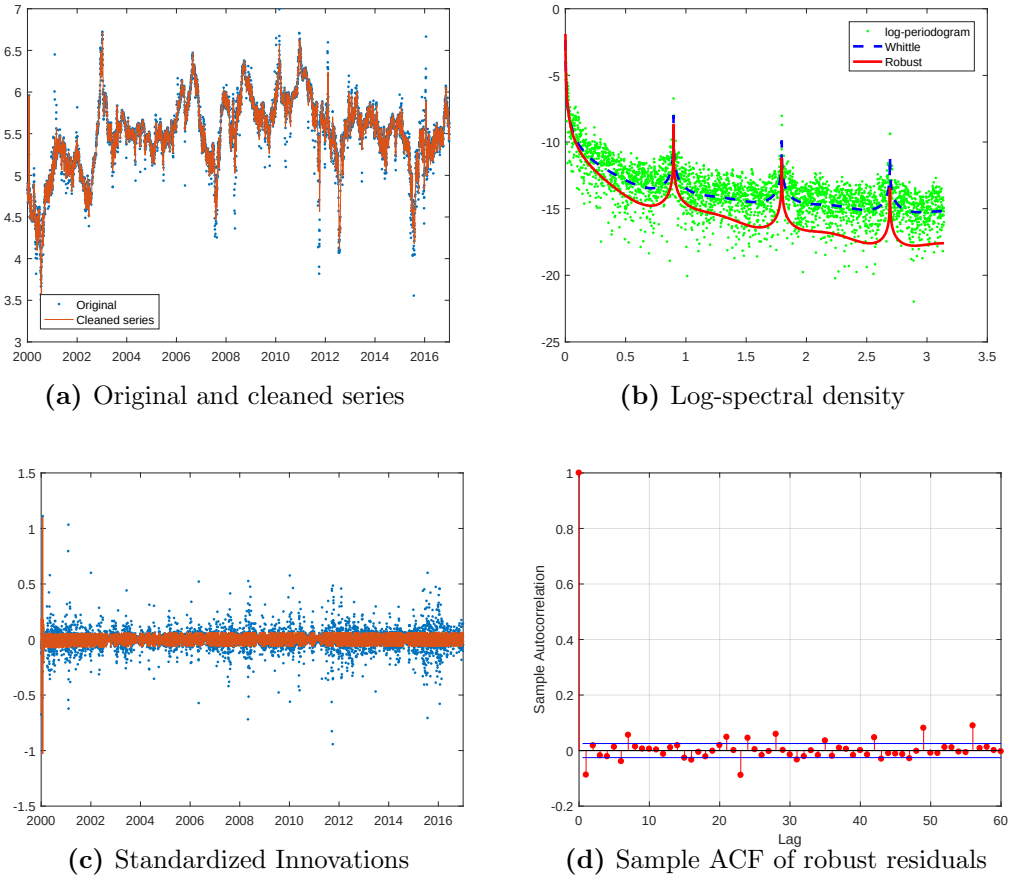
Table 1: Nord Pool daily average prices (logs), 1/1/2000-31/12/2016. Point estimates and estimation standard errors of the memory a short run cepstral parameters obtained by Whittle maximum likelihood estimation, robust estimation (one iteration), iterative robust estimation. The tuning parameter for robust estimation is set equal to $a = 1.345$.

	Whittle		Robust 1st iter		Robust final	
Coeff.	Point Est	StdErr	Point Est	StdErr	Point Est	StdErr
d_0	0.8336	0.0370	0.9197	0.0370	0.9137	0.0370
d_1	0.4170	0.0258	0.4520	0.0258	0.4561	0.0258
d_2	0.3813	0.0261	0.4249	0.0261	0.4373	0.0261
d_3	0.3612	0.0258	0.3571	0.0258	0.3833	0.0258
c_{z0}	-13.7167	0.0179	-15.2010	0.0179	-15.3960	0.0179
c_{z1}	0.2934	0.0641	0.4718	0.0641	0.6824	0.0641
c_{z2}	0.1231	0.0349	0.2130	0.0349	0.1620	0.0349
c_{z3}	0.1484	0.0253	0.1555	0.0253	0.1612	0.0253
c_{z4}	0.0817	0.0208	0.1164	0.0208	0.1161	0.0208
c_{z5}	0.0568	0.0182	0.1102	0.0182	0.1110	0.0182
c_{z6}	0.0690	0.0164	0.1292	0.0164	0.1545	0.0164
c_{z7}	-0.1386	0.0204	-0.1444	0.0204	-0.1663	0.0204

The most noticeable fact is that the robustified estimates of the long memory parameter at the zero frequency increases from 0.83 to 0.92. The memory of the fundamental and harmonic weekly cycle also increse, expect perhaps for d_3, although the robust estimates remain within the stationary region. The prediction error variance, which is estimated by taking the exponential of $\hat{c}_0$, is smaller for the robust case. Also, iterating the robust estimation procedure, does not change the estimates relevantly. Notice that the asymptotic standard errors do not vary, as they are a function of the explanatory variables in the logarithmic spectral model 3.

Figure 2a compares the original series and the cleaned series arising from the iterative robust estimate. Figure 2b compares the estimated log-spectra. It is evident that the robust estimate diverges more rapidly around the zero and seasonal frequencies, and that the log-spectrum is smaller that the one estimated on the original series, as the data-cleaning filter has reduced the contribution of the noise to the spectrum. This is clearly visible from the plot of the standardized innovations $\nu_t/\sqrt{f_t}$ for the original series and $\nu_t^+/\sqrt{f_t^+}$ for the clean series, in figure 2c. Figure 2d plots the sample autocorrelation function of the robust standardized residuals. As it may be anticipated from these plots, the presence of conditional heteroscedastity in the robust time series residuals is less of a problem.

As a matter of fact, some residual persistence has ended up in the spikes estimates. The average value of a spike (defined as y_t minus its cleaned estimate) is close to zero, but there is some degree of clustering: in particular, the first eight sample autocorrelations of the spike sequence are:



(a) Original and cleaned series **(b)** Log-spectral density

(c) Standardized Innovations **(d)** Sample ACF of robust residuals

Figure 2: Nord Pool average daily prices (logs). Original and cleaned series (2a); estimated log-spectrum (2b); standardized innovations (2d) and their sample autocorrelation function (2d).

lag 1	lag 2	lag 3	lag 4	lag 5	lag 6	lag 7	lag 8
0.5458	0.3229	0.2591	0.1848	0.1105	0.1017	0.1317	-0.0143

We now turn to the issue of choosing the tuning parameter. The value 1.345 is a popular choice in the robustness literature, but, as we saw above, it presumably oversmooths the series, delivering a cleaned series which is less informative about the future than the original one. The optimal value of a will be determined by time series cross-validation, by looking at the ability to predict electricity prices h steps ahead, where h ranges from 1 day to 14 days.

For this purpose we perform a rolling forecast experiment such that, starting from the first of January 2015, we estimate the model by Whittle likelihood on the original and the cleaned data (from a single iteration of the algorithm) from 1/1/2000 to 31/12/2014, using values of a ranging from 1 to 5 with step 1/3. Conditional on the parameter estimates, we compute the out-of-sample predictions $\tilde{y}_{n_1+h|n_1}$, for $h = 1, \ldots, 14$ and $n_1 = 5,479$. This yields 13 sets of predictions, corresponding to the different values of a. Subsequently, we update the training sample by adding the observation $n_1 + 1$ and removing the first, and perform the same operations until the end of sample is reached. Comparing the prediction

to the actual observations yields $731 - h$ h-step-ahead forecast errors that can be used for assessing the predictive accuracy.

Figure 3 is a contour plot of the logarithm relative mean square forecast error of the robust predictor, versus the raw predictor, as a function of h and a, denoted RelMSE(h, a). Values smaller than zero are such that the robust predictor outperforms the raw one. As

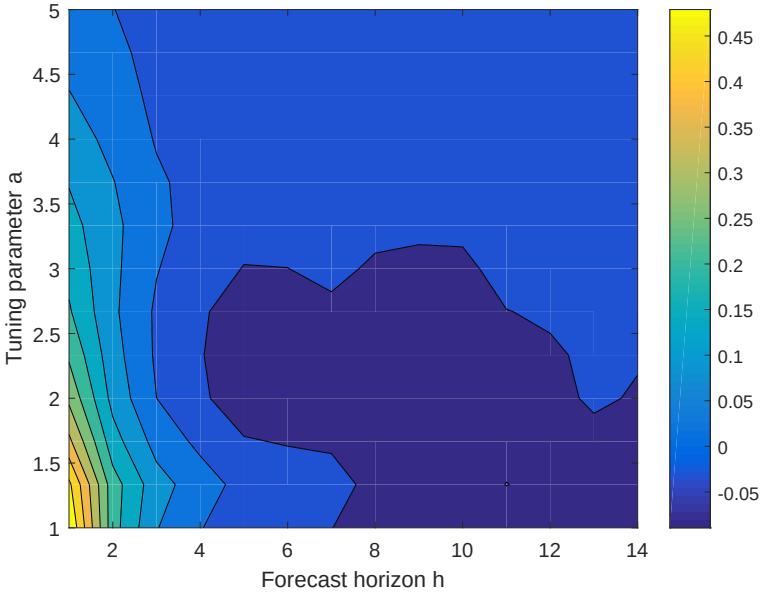


Figure 3: Logarithm of the relative mean square forecast error of the robust predictor as a function of the forecast horizon $h = 1, \ldots, 14$ (horizontal axis) and the tuning parameter a (vertical axis).

far as day-ahead prediction ($h = 1$) is concerned, robustification does not provide an advantage: the relative MSE as a varies converges towards 1 and we need to have a large value of a to obtain the same performance, as the following table shows:

a	1	2	3	4	5
RelMSE$(1, a)$	1.71	1.34	1.20	1.12	1.04

This seemingly disappointing result might be due to the fact that spikes appear in short clusters, affecting neighbouring observations. However, already for $h = 4$ the robust predictor is more accurate if we choose a around 2.33, which is higher than the default value 1.33. More generally, the robust predictor can outperform the usual predictor for longer horizons. In the plot, the light blue area corresponds to the combinations which yield equal predictive accuracy as the predictor based on the original data. The dark blue area is such that the largest forecasting accuracy gains accrue from robustifying the estimates.

The results of the rolling forecasting experiment point out that the extent of the robustification that is needed to optimize the out-of-sample predictive ability of the maintained model is less than that implied by the threshold $a = 1.345$, i.e. it is achieved for larger values of a. This conclusion is reinforced by the analysis of the standardized residuals of

the model estimated using the full sample, according to the robust estimation methodology outlined in section 4, using different values of a in the range [1,5]. The following table presents the value of the skewness, kurtosis and Jarque and Bera (1980) statistics.

	Values of a									
	1.00	1.33	1.67	2.00	2.33	2.67	3.00	4.00	5.00	∞
Skewness	-0.12	-0.05	-0.03	-0.03	-0.03	-0.02	-0.03	-0.01	-0.02	-0.06
Kurtosis	1.68	1.93	2.18	2.43	2.71	3.01	3.31	4.30	5.36	18.15
Jarque-Bera	463.90	298.10	175.34	84.61	22.63	0.64	25.49	431.39	1433.65	$5.89 \cdot 10^4$

The last column refers to the standard non robust case, such that the standardized residuals are computed by the usual KF (see section 3.2) conditional on the parameter estimates maximising the Whittle likelihood. The minimum of the Jarque-Bera statistic is obtained for a around 2.67. Thus, it is likely that the default value for the tuning constant results in a loss of the informational content of the observed data.

6 Conclusions

Price spikes affect both the parameter estimates and the predictions arising from a model. With respect to the predictions, they do not only exert their influence by biasing downwards the estimates of the persistence of the system's dynamics, but their effect is also propagated via the forecast function, if we condition on their observed values.

This paper has provided a correction for those distorting effects based on a "data cleaning" filter, which modifies the basic Kalman filter equation in a way that a suspicious observation is corrected by shrinking it towards its expected value, given the past observations. The identification of a spike relies on the selection of a threshold defined on the scale of the time series innovations. We have addressed the choice of the optimal level of robustification by choosing the threshold according to criterion of maximising the out-of-sample predictability.

Our main conclusion is that robustifying the parameter estimates and the predictions contributes positively to the forecasting accuracy of the maintained model at longer leads; furthermore, the forecasting performance may be optimized with respect to the choice of the tuning parameter. At shorter horizons, such as one day ahead, we found that no gains accrue from robust estimation and prediction. This result may be a consequence of the fact that spikes arrive in short clusters, so that a spiky observations at time t contains important information for predicting next day's price.

References

Beran, J., Feng, Y., Ghosh, S., and Kulik, R. (2013). *Long-memory processes: Probabilistic properties and statistical methods*. Berlin, Springer.

Bloomfield, P. (1973). An exponential model for the spectrum of a scalar time series. *Biometrika*, *60*, 217-226.

Bogert, B., Healy, M., and Tukey, J. (1963). The frequency analysis of time series for echoes: Spectrum, pseudo-autocovariance, cross-spectrum and shape cracking. *In Proc. Symp. on Time Series Analysis*, *1*, 209-243.

Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *The Annals of Statistics*, *17*, 1749-1766.

Dupuis, D. J. (2017). Electricity price dependence in new york state zones: A robust detrended correlation approach. *The Annals of Applied Statistics*, *11*(1), 248–273.

Giraitis, L., Koul, H., and Surgailis, D. (2012). *Large sample inference for long memory processes* (Tech. Rep.). Imperial College Press.

Gradshteyn, I., and Ryzhik, I. (2007). *Table of integrals, series, and products*. Amsterdam, Seventh edition: Elsevier/Academic Press.

Gray, L., Zhang, N., and Woodward, A. (1989). On generalized fractional processes. *Journal of Time Series Analysis*, *10*, 233–257.

Hellström, J., Lundgren, J., and Yu, H. (2012). Why do electricity prices jump? empirical evidence from the nordic electricity market. *Energy Economics*, *34*(6), 1774–1781.

Hosking, J. (1981). Fractional differencing. *Biometrika*, *68*, 165-176.

Hsu, N., and Tsai, H. (2009). Semiparametric estimation for seasonal long-memory time series using generalized exponential models. *Journal of Statistical Planning and Inference*, *139*, 1992-2009.

Hurvich, C. M. (2002). Multistep forecasting of long memory series using fractional exponential models. *International Journal of Forecasting*, *18*, 167–179.

Jancek, G. (1982). Determining the degree of differencing for time series via the log spectrum. *Journal of Time Series Analysis*, *3*, 177-183.

Jarque, C. M., and Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics letters*, *6*(3), 255–259.

Maronna, R., Martin, D., and Yohai, V. (2006). *Robust Statistics*. John Wiley & Sons, Chichester. ISBN.

Martin, R., and Thomson, D. (1982). Robust-resistant spectrum estimation. *Proceedings of the IEEE*, *70*, 1097–1115.

Masreliez, C., and Martin, R. (1977). Robust bayesian estimation for the linear model and robusting the kalman filter. *IEEE Transactions of Automatic Control*, 361–371.

McCloskey, A., and Hill, J. (2013). *Parameter estimation robust to low-frequency contamination*. Journal of Business & Economic Statistics.

NordPool. (2017). *Nord pool spot*. Retrieved 2017-07-15, from `http://www.nordpoolspot.com/`

Nowotarski, J., Tomczyk, J., and Weron, R. (2013). Robust estimation and forecasting of the long-term seasonal component of electricity spot prices. *Energy Economics*, *39*, 13–27.

Pearlman, J. G. (1980). An algorithm for the exact likelihood of a high-order autoregressive-moving average process. *Biometrika*, *67*, 232–233.

Pourahmadi, M. (1983). Exact factorization of the spectral density and its application to forecasting and time series analysis. *Communications in Statistics - Theory and Methods*, *12*, 2085–2094.

Soares, L., and Souza, L. (2006). Forecasting electricity demand using generalized long memory. *International Journal of Forecasting*, *22*, 17–28.

Velasco, C. (1995). Non-stationary log-periodogram regression. *Journal of Econometrics*, *91*, 325–371.

Velasco, C., and Robinson, P. (2000). Whittle pseudo-maximum likelihood estimation for nonstationary time series. *Journal of the American Statistical Association*, *95*, 1229–1243.

Woodward, W., Gray, H., and Elliott, A. (2011). *Applied time series analysis* (1edition ed.). CRC Press.

Research Papers 2017

2017-23: Roman Frydman, Søren Johansen, Anders Rahbek and Morten Nyboe Tabor: The Qualitative Expectations Hypothesis: Model Ambiguity, Concistent Representations of Market Forecasts, and Sentiment

2017-24: Giorgio Mirone: Inference from the futures: ranking the noise cancelling accuracy of realized measures

2017-25: Massimiliano Caporin, Gisle J. Natvik, Francesco Ravazzolo and Paolo Santucci de Magistris: The Bank-Sovereign Nexus: Evidence from a non-Bailout Episode

2017-26: Mikkel Bennedsen, Asger Lunde and Mikko S. Pakkanen: Decoupling the short- and long-term behavior of stochastic volatility

2017-27: Martin M. Andreasen, Jens H.E. Christensen and Simon Riddell: The TIPS Liquidity Premium

2017-28: Annastiina Silvennoinen and Timo Teräsvirta: Consistency and asymptotic normality of maximum likelihood estimators of a multiplicative time-varying smooth transition correlation GARCH model

2017-29: Cristina Amado, Annastiina Silvennoinen and Timo Teräsvirta: Modelling and forecasting WIG20 daily returns

2017-30: Kim Christensen, Ulrich Hounyo and Mark Podolskij: Is the diurnal pattern sufficient to explain the intraday variation in volatility? A nonparametric assessment

2017-31: Martin M. Andreasen, Jens H.E. Christensen and Glenn D. Rudebusch: Term Structure Analysis with Big Data

2017-32: Timo Teräsvirta: Nonlinear models in macroeconometrics

2017-33: Isabel Casas, Eva Ferreira and Susan Orbe: Time-varying coefficient estimation in SURE models. Application to portfolio management

2017-34: Hossein Asgharian, Charlotte Christiansen, Ai Jun Hou and Weining Wang: Long- and Short-Run Components of Factor Betas: Implications for Equity Pricing

2017-35: Juan Carlos Parra-Alvarez, Olaf Posch and Mu-Chun Wang: Identification and estimation of heterogeneous agent models: A likelihood approach

2017-36: Andrés González, Timo Teräsvirta, Dick van Dijk and Yukai Yang: Panel Smooth Transition Regression Models

2017-37: Søren Johansen and Morten Ørregaard Nielsen: Testing the CVAR in the fractional CVAR model

2017-38: Nektarios Aslanidis and Charlotte Christiansen: Flight to Safety from European Stock Markets

2017-39: Tommaso Proietti, Niels Haldrup and Oskar Knapik: Spikes and memory in (Nord Pool) electricity price spot prices