

# Prediction interval of electricity prices by robust nonlinear models

Lisa Crosato, Luigi Grossi and Fany Nan.

**Abstract** It is well known that volatility of electricity prices estimated through GARCH-type models can be strongly affected by the presence of extreme observations. Although the presence of spikes is a well-known stylized effect observed on electricity markets, their presence has been often neglected and robust estimators have been rarely applied. In this paper we try to fill this gap introducing a robust procedure to the study of the dynamics of electricity prices. The conditional mean of de-trended and seasonally adjusted prices is modeled through a robust estimator of SETAR processes based on a polynomial weighting function (Grossi and Nan, 2015), while a robust GARCH is used for the conditional variance. The robust GARCH estimator relies on the extension of the forward search by Crosato and Grossi (2017). The robust SETAR-GARCH model is applied to the Italian electricity markets using data in the period spanning from 2013 to 2015. The purpose of this application is therefore twice: first, it is possible to enhance the prediction from point to intervals with associated probability levels, second, we set up a procedure to detect possible extreme prices which are commonly observed in electricity markets.

## 1 Introduction

In this paper we introduce a doubly robust approach to modelling the volatility of electricity spot prices, minimizing the misleading effects of the extreme jumps that characterize this particular kind of data on the predictions. The purpose of this paper is twice: first, it is possible to enhance the prediction from point to intervals with associated probability levels, second, we can detect possible extreme prices which are commonly observed in electricity markets. Although many papers have applied quite sophisticated time series models to prices and demand time series of electricity and gas very few have considered the strong influence of jumps on estimates and the need to move to robust estimators (Nowotarski et al.; 2013). Thus, the big differences between our paper and the previous literature related to jumps

---

Lisa Crosato  
University of Milano-Bicocca, Via Bicocca degli Arcimboldi 8, 20126 Milano. e-mail: lisa.crosato@unimib.it.

Luigi Grossi  
University of Verona, Via Cantarane, 24, 37129 Verona. e-mail: luigi.grossi@univr.it.

Fany Nan  
Joint Research Center of the European Commission, Ispra. e-mail: fany.nan@ec.europa.eu.

in electricity prices are twofold: 1) we don't detect extreme observations and we don't focus on the prediction of jumps; 2) We apply robust estimators which are not strongly influenced by the jumps. In this way we improve the prediction of normal prices which represent the majority of data. Our approach could then be integrated with other methods dealing with jumps forecasting.

Our methodology is applied separately to the hourly time series of Italian electricity price data from January 1st, 2013 to December 31th, 2015. In the first step, we apply a threshold autoregressive model (SETARX) to the time series of logarithmic prices. In the second step the weighted forward search estimator (WFSE) for GARCH(1,1) models is applied to the residuals extracted from the first step in order to estimate and forecast volatility. The weighted forward search is a modification of the Forward Search intending to correct the effects of extreme observations on the estimates of GARCH(1,1) models. Differently from the original Forward Search, at each step of the search estimation involves all observations which are weighted according to their degree of outlyingness.

## 2 A robust SETAR model for electricity prices

It is well-known that the series of electricity prices have long-run behaviour and annual dynamics, which change according with the load period. A common characteristic of price time series is the weekly periodic component (of period 7), suggested by the spectra that show three peaks at the frequencies  $1/7$ ,  $2/7$  and  $3/7$ , and a very persistent autocorrelation function.

We assume that the dynamics of log prices can be represented by a nonstationary level component  $L_{tj}$ , accounting for level changes and/or long-term behaviour, and a residual stationary component  $p_{tj}$ , formally:

$$\log P_{tj} = L_{tj} + p_{tj}. \quad (1)$$

To estimate  $L_{tj}$  we use the wavelets approach (Lisi and Nan, 2014). We consider the Daubechies least asymmetric wavelet family, LA(8), and the coefficients were estimated *via* the maximal overlap discrete wavelet transform (MODWT) method (for details, see Percival and Walden, 2000).

The detrended prices  $p_t$ , will then be modeled by a two-regime Self-Exciting Threshold AutoRegressive model SETAR(7,1) which is specified as

$$p_t = \begin{cases} \mathbf{x}_t \boldsymbol{\beta}_1 + \varepsilon_t, & \text{if } p_{t-1} \leq \gamma \mathbf{x}_t \boldsymbol{\beta}_2 + \varepsilon_t \\ \gamma \mathbf{x}_t \boldsymbol{\beta}_2 + \varepsilon_t & \text{if } p_{t-1} > \gamma \end{cases} \quad (2)$$

for  $t = 1, \dots, N$ , where  $p_{t-1}$  is the threshold variable and  $\gamma$  is the threshold value. The relation between  $p_{t-1}$  and  $\gamma$  states if  $p_t$  is observed in regime 1 or 2.  $\boldsymbol{\beta}_j$  is the parameter vector for regime  $j = 1, 2$  containing 7 coefficients to account for weekly periodicity.  $\mathbf{x}_t$  is the  $t$ -th row of the  $(N \times 7)$  matrix  $\mathbf{X}$  comprising 7 lagged variables of  $p_t$  (and eventually a constant). Errors  $\varepsilon_t$  are assumed to follow an iid(0,  $\sigma_\varepsilon$ ) distribution.

Parameters can be estimated by sequential conditional least squares. In the case of robust two-regime SETAR model, for a fixed threshold  $\gamma$  the GM estimate of the autoregressive parameters can be obtained by applying the iterative weighted least squares:

$$\hat{\beta}_j^{(n+1)} = \left( \mathbf{X}_j' \mathbf{W}^{(n)} \mathbf{X}_j \right)^{-1} \mathbf{X}_j' \mathbf{W}^{(n)} \mathbf{p}_j \quad (3)$$

where  $\hat{\beta}_j^{(n+1)}$  is the GM estimate for the parameter vector in regime  $j = 1, 2$  after the  $n$ -th iteration from an initial estimate  $\hat{\beta}_j^{(0)}$ , and  $\mathbf{W}^{(n)}$  is a weight diagonal matrix, those elements depend on a weighting function  $w(\hat{\beta}_j^{(n)}, \hat{\sigma}_{\varepsilon,j}^{(n)})$  bounded between 0 and 1. The threshold  $\gamma$  can be estimate by minimizing the objective function  $\rho(r_t)$  over the set  $\Gamma$  of allowable threshold values. Weights are calculated as

$$w(\hat{\beta}_j, \hat{\sigma}_{\varepsilon,j}) = \psi \left( \frac{y_t - m_{y,j}}{C_y \hat{\sigma}_{y,j}} \right) \psi \left( \frac{y_t - \mathbf{x}_t \hat{\beta}_j}{C_\varepsilon \hat{\sigma}_{\varepsilon,j}} \right)$$

where  $\psi$  is the polynomial weight function (Maronna et al., 2006) and  $m_{y,j}$  is a robust estimate of the location parameter (sample median) in the  $j$ -th regime.  $\hat{\sigma}_{y,j}$  and  $\hat{\sigma}_{\varepsilon,j}$  are robust estimates of the scale parameters  $\sigma_y$  and  $\sigma_\varepsilon$  respectively, obtained by the median absolute deviation multiplied by 1.483.  $C_y$  and  $C_\varepsilon$  are tuning constants fixed at 6.0 and 3.9 respectively.

### 3 Robust volatility estimation through the forward search

Now we apply the weighted forward search (WFS) estimator to derive robust prediction intervals for the volatility of electricity prices, starting from the residuals of the SETAR model estimated in section 1. Let  $\varepsilon_t$  denote an observed time series of heteroscedastic residuals. For electricity prices  $\varepsilon_t = p_t - \hat{p}_t$  where  $\hat{p}_t$  is the price fitted by the robust SETAR model. We proceed now by estimating a standard GARCH(1,1) model on residuals  $\varepsilon_t$ , so that  $\varepsilon_t | F_{t-1} \sim N(0, \sigma_t^2)$  and

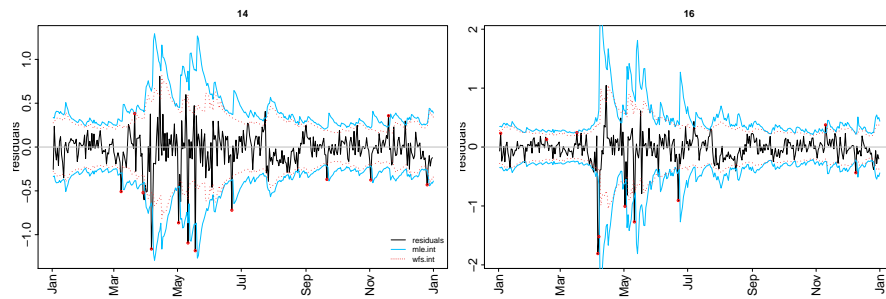
$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \quad (4)$$

with  $\alpha_0 > 0, \alpha_1 \geq 0, \beta \geq 0, \alpha_1 + \beta < 1$ .

The WFS (see Crosato and Grossi, 2017, for details) starts with the selection of a small subset of observations showing the minimum median of squared residuals. Then, at each step of the search, MLE estimation is carried out on all observations weighted to account for their degree of outlyingness so that the estimation pattern along the search is not influenced by the presence of spikes. All units, a part of the first  $b$ , are ranked according to their squared standardized residuals with respect to the model estimated at the previous step of the search. Proceeding along the search, an increasing number of units are given weight 1 while the remaining are down-weighted by the corresponding value of the complementary cumulative distribution

function of the squared standardized residuals defined on the whole sample. The weights range from approximately 0 to 1 so that outliers will be heavily down-weighted until the end of the search while the closer the weight to 1 the stronger is the degree of agreement of the observation with the estimated GARCH. This way the temporal structure of the time series is respected, filling the time gaps created by the forward search ordering but on the same time observations are ordered consistently with the GARCH model estimated until the last steps when all observations enter with their original value.

A plus of the WFS approach is that outliers are identified automatically through a test reducing the arbitrariness in declaring a given observation as outlier. Once outliers have been identified the forward plots of coefficient estimates are cut automatically and WFS GARCH estimates are obtained.



**Fig. 1** Forecasted conditional confidence intervals of volatility for hourly electricity prices in Italy (hour 14 and 16, January to December 2015) obtained by MLE (light blue) and WFS (dotted red).

The outliers identified through the WFS test vary from a minimum of 29 (hour 21) to a maximum of 59 (hours 8 and 9). A few outliers characterize many hours of the same day, as is the case of the 21st and 27th of April, 2013 for 14 hours as well as hot days of June 2013. Spikes in 2014 characterize mainly the months of February, March and April. Outliers for 2015, hours 14 and 16, are highlighted by red circles in figure 1.

Using the MLE and WFS estimated coefficients and applying equation (4) for the conditional variance recursively, we then obtain the forecasted price volatility and the corresponding 95 percent confidence intervals (see figure 1). As can be seen, the WFS intervals (dotted red) are tighter around the realized volatility than the MLE ones (light blue). The gain in prediction is evident in the aftermath of a spike or a drop, for instance in April and May. Note that the correction provided by the WFS works also for smallest jumps, as in July and December.

## References

1. Crosato L., Grossi L.: Correcting outliers in GARCH models: a weighted forward approach. *Statistical Papers* (2017) doi:10.1007/s00362-017-0903-y.
2. Grossi L., F. Nan: Robust estimation of regime switching models. In: Morlini I., Minerva T., Palumbo F. (eds.) *Advances in Statistical Models for Data Analysis*, pp. 125–135. Springer International Publishing, Switzerland (2015).

3. Lisi, F., Nan, F.: Component estimation for electricity prices: Procedures and comparisons. *Energy Economics* **44**, 143–159 (2014).
4. Nowotarski J., Tomczyk J., Weron R.: Robust estimation and forecasting of the long-term seasonal component of electricity spot prices. *Energy Economics* **39**, 13–27 (2013).
5. Percival, D., Walden, A.: *Wavelet Methods for Time Series Analysis*. Cambridge University Press. (2000).