Working Paper Series
Department of Economics
University of Verona

# Forecasting electricity prices through robust nonlinear models

Luigi Grossi, Fany Nan

WP Number: 6 May 2017

ISSN: 2036-2919 (paper), 2036-4679 (online)

# Forecasting electricity prices through robust nonlinear models*

Luigi Grossi[a], Fany Nan[b]

[a]*University of Verona, Department of Economics, Verona, Italy*
[b]*European Commission, Joint Research Center, Ispra, Italy*

**Abstract**

In this paper a robust approach to modelling electricity spot prices is introduced. Differently from what has been recently done in the literature on electricity price forecasting, where the attention has been mainly drawn by the prediction of spikes, the focus of this contribution is on the robust estimation of nonlinear SETARX models. In this way, parameters estimates are not, or very lightly, influenced by the presence of extreme observations and the large majority of prices, which are not spikes, could be better forecasted. A Monte Carlo study is carried out in order to select the best weighting function for GM-estimators of SETAR processes. A robust procedure to select and estimate nonlinear processes for electricity prices is introduced, including robust tests for stationarity and nonlinearity and robust information criteria. The application of the procedure to the Italian electricity market reveals the forecasting superiority of the robust GM-estimator based on the polynomial weighting function on the non-robust Least Squares estimator. Finally, the introduction of external regressors in the robust estimation of SETARX processes contributes to the improvement of the forecasting ability of the model.

*Keywords:* Electricity price, Nonlinear time series, Price forecasting, Robust GM-stimator, Spikes, Threshold models

*JEL codes*: C13, ,C15, C22, C53, Q47

*The views expressed are purely those of the authors and may not in any circumstances be regarded as stating an official position of the European Commission.

*May 8, 2017*

## 1. Introduction

Forecasting electricity prices is a crucial objective for many reasons (Nogales et al., 2002). First of all, speculative trading on electricity markets has become more and more important, especially on the short run. Strictly related to trading is the possibility to evaluate the economic convenience of short-run electricity storage facilities which would be of great importance for the strategic role they could play on the integration of intermittent renewable sources into the grid (Flatley et al., 2016). From the regulator perspective, it is of vital relevance the ability to predict future prices in order to reduce the risk of volatility and its impact on final consumers (Hong et al., 2016). Also generators are interested in future prices for driving the decision related to the capacity size of the plants and to the load to produce and inject into the grid (Aggarwal et al., 2009). With an accurate day-ahead price forecast, a producer can develop an appropriate bidding strategy to maximize ones own benefit, or a consumer can maximize its utility (Conejo et al., 2005). For a very detailed discussion of the relevance of electricity price forecasting, see Weron (2014).

Spot electricity prices are known to exhibit sudden and very large jumps to extreme levels as a consequence of sudden grid congestions, unexpected shortfalls in supply, and failures of the transmission infrastructure (Christensen et al., 2012). Such events reflect immediately on prices because of the non-storable nature of electrical energy and the requirement of a constant balance between demand and supply (Huisman & Mahieu, 2003). This feature must be considered very carefully and robust techniques must be applied to avoid that few jumps could dramatically affect parameter estimates and, consequently, forecasts.

Several papers have dealt with the issue of modelling spikes in electricity prices. Particularly used have been diffusion processes introducing spikes through the addition of a Poisson jump component (Cartea & Figueroa, 2005; Escribano et al., 2011). Processes with heavy-tailed distribution have instead been estimated by Bystrom (2005), Panagiotelis & Smith (2008) and Swider & Weber (2007). Other authors have coped with the issue of predicting price spikes which are particularly relevant for risk management (Laouafi et al., 2016). In this context, Christensen et al. (2012) suggested a modified autoregressive conditional hazard model to predict price spikes on the Australian electricity market. Clements et al. (2013) proposed a semi-parametric model for price spikes forecasting. The necessity to resort to nonlinear time series models has been pointed out, among others, by Bordignon et al. (2013) where Markov switching models are applied to forecast prices on the UK electricity market. Other authors have applied threshold autoregressive models (Ricky Rambharat et al., 2005; Zachmann, 2013; Haldrup & Nielsen, 2006; Lucheroni, 2012; Sapio & Spagnolo, 2016) to separate a normal regime, when volatility is rather low, and a high volatility regime when spikes are observed. The superiority of regime switching models with respect to models without regimes has been argued by Janczura & Weron (2010) and Kosater & Mosler (2006), who have observed better forecasting performances for nonlinear processes. An interesting approach has been suggested recently by Gaillard et al. (2016), who predict the maximal price of the day, which is then used as an exogenous variable in a prediction model based on a quantile regression estimator.

Although many papers have applied quite sophisticated time series models to time series of electricity and gas prices and demand with spikes, only few have considered the strong influence of jumps on estimates and the need to move to robust estimators (Janczura et al., 2013; Nowotarski et al., 2013; Haldrup et al., 2016).

In the present paper we suggest to use a version of threshold autoregressive models (SETARX) where parameters are estimated robustly to the presence of spikes. Differently from what has been done in the literature so far, we are not interested in modelling spikes, but we want to focus the attention on the influence that spikes can have on the estimated coefficients. If non robust estimators are applied, coefficient could be very badly biased and even non-spiky observations, which are the very large majority, could not be properly modeled and forecasted.

Moreover, we suggest a completely robust approach to modelling and forecasting electricity prices which embed robust estimation of a SETARX model, robust tests for unit root and nonlinear components and robust information criteria. Although we are aware of the limits of this class of models (Misiorek et al., 2006), threshold models represent a simple approach which takes into account the possible nonlinearity of electricity prices and allows the inclusion of external regressors to improve their forecasting performances (Maciejowska et al., 2016).

Threshold Auto Regressive (TAR) models are quite popular in the nonlinear time-series literature. This popularity is due to the fact that they are relatively simple to specify, estimate, and interpret. The sampling properties of the estimators and test statistics associated with TAR models have been studied by Tsay (1989) and Hansen (1997, 1999). In the class of non-linear models, studies addressed to robustifying this kind of models are very few, although the problem is very challenging, particularly when it is not clear whether aberrant observations must be considered as outliers or as generated by a real non-linear process. van Dijk (1999) derived an outlier robust estimation method for the parameters in Smooth Threshold Auto Regressive (STAR) model, based on the principle of generalized maximum likelihood type estimation. Battaglia & Orfei (2005) focused on outlier detection and estimation through a model-based approach when the time series is generated by a general non-linear process. A general model able to capture nonlinearity, structural changes and outliers has been introduced by Giordani et al. (2007). The authors suggest to employ the state-space framework which allows to estimate the coefficients of several non-linear time series models and simultaneously take into account the presence of outliers and structural breaks. The method seems quite effective in modeling macro-economic time series.

Apart from the previous methods which deal with the presence of outliers in very specific contexts, the issue of outliers in non-linear time series models is far from being clearly solved. Chan & Cheung (1994) extended the generalized M estimator method[1] to Self-Exciting Threshold Auto Regressive (SETAR) models. Their simulation results show that the GM estimation is preferable to the LS estimation in presence of additive outliers. As GM estimators have proved to be consistent with a very small loss of efficiency, at least under normal assumptions, the extension to threshold models, which are piecewise linear, looks quite straightforward. Despite this observation, a cautionary note has been written by Giordani (2006) to point out some drawbacks of the GM estimator proposed by Chan & Cheung (1994). In particular, it is argued and shown, by means of a simulation study, that the GM estimator can deliver inconsistent estimates of the threshold even under regularity conditions. According to this contribution, the inconsistency of the estimates could be particularly severe when strongly descending weight functions are used.

Zhang et al. (2009) demonstrate the consistency of GM estimators of autoregressive parameters in each regime of SETAR models when the threshold is unknown. The consistency of parameters is guaranteed when the objective function is a convex non-negative function. A possible function holding these properties is the Huber $\rho-$function which is suggested to replace the polynomial function used in Giordani's (2006) paper. However, the authors conclude, the problem of finding a threshold robust estimator with desirable finite-sample properties is still an open issue. Although, a theoretical proof has been provided by the authors, there is not a well structured Monte Carlo

[1]For an overview about GM estimators see (Andersen, 2008, chap. 4) and (Maronna et al., 2006, chap. 8.5)

study to assess the extent of the distortion of the GM-SETAR estimator. Thus, from the analysis of the existing literature, it is not clear the extent of the bias of robust estimators of the threshold with respect to LS estimator, how to choose the best weighting function and the forecasting performances of different weighting functions have never been compared. Moreover, robust estimators of regime switching processes are not implemented within the most popular software platforms among statisticians, such as Matlab and R.

Grossi & Nan (2015) have started to address the above points through a Monte Carlo experiment which compares the performances of classical SETAR estimator and robust estimator using different weighting functions. The main insights obtained from that preliminary work are confirmed in the present paper where a more extensive simulation experiment is carried out. The simulation experiment has required the implementation of all the estimators (classical and robust) in R language resulting in a set of functions which hopefully will become a library soon.

The results obtained from the simulation experiment are used to estimate the parameters of SETAR models on the Italian electricity price data (*PUN, prezzo unico nazionale*). The model is enriched by the introduction of exogenous regressors which improve the forecasting performances. Crucial variables in predicting electricity prices are dummies for the intra-weekly seasonality, predicted demanded volumes and predicted wind power generation (Gianfreda & Grossi, 2012).

Summarizing, the main contributions of the present paper are:

- a Monte Carlo simulation study is performed to integrate partial simulations done in previous papers. At the end of this study the best robust estimator is clearly detected;
- a robust approach to modelling and forecasting electricity prices is suggested which include tests, estimation of parameters and selection of the best model;
- a robust nonlinear model with exogenous regressors is estimated which takes into account the main stylized facts observed on electricity markets and includes the forecasted regressors which have revealed to increase substantially the forecasting performances (Gaillard et al., 2016; Weron, 2014).

The structure of the paper is as follows. In section 2 the general SETAR model is defined and different weighting functions used to robustify the classic estimator are discussed. Section 3 contains the main results of the Monte Carlo simulation study. The analysis of the forecasting performances of the robust SETARX model based on the polynomial weighting function is presented in section 4. Section 5 reports some concluding remarks and suggestions for future research.

## 2. SETAR models with exogenous regressors

Given a time series $y_t$, a two-regime Self-Exciting Threshold Auto Regressive model SETAR($p$,$d$) with exogenous regressors is specified as

$$y_t = \begin{cases} \mathbf{x}_t\boldsymbol{\beta}_1 + \mathbf{z}_t\boldsymbol{\lambda}_1 + \varepsilon_{1t}, & \text{if} \quad y_{t-d} \leq \gamma \\ \mathbf{x}_t\boldsymbol{\beta}_2 + \mathbf{z}_t\boldsymbol{\lambda}_2 + \varepsilon_{2t}, & \text{if} \quad y_{t-d} > \gamma \end{cases} \tag{1}$$

for $t = \max(p, d), ..., N$, where $y_{t-d}$ is the threshold variable with $d \geq 1$ and $\gamma$ is the threshold value. The relation between $y_{t-d}$ and $\gamma$ states if $y_t$ is observed in regime 1 or 2. $\boldsymbol{\beta}_j$ is the vector of auto-regressive parameters for regime $j = 1, 2$ and $\mathbf{x}_t$ is the $t$-th row of the $(N \times p)$ matrix $\mathbf{X}$ comprising $p$ lagged variables of $y_t$. $\boldsymbol{\lambda}_j$ is the vector of parameters corresponding to exogenous regressors and/or dummies contained in the $(N \times r)$ matrix $\mathbf{Z}$ whose $t$-th row is $\mathbf{z}_t$. Errors $\varepsilon_{1t}$ and $\varepsilon_{2t}$ are assumed to follow distributions $\text{iid}(0, \sigma_{\varepsilon,1})$ and $\text{iid}(0, \sigma_{\varepsilon,2})$ respectively.

### *2.1. Estimation of SETAR models*

In general the value of the threshold $\gamma$ is unknown, so that the parameters to estimate become $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}_1', \lambda_1')'$, $\boldsymbol{\theta}_2 = (\boldsymbol{\beta}_2', \lambda_2')'$, $\gamma$, $\sigma_{\varepsilon,1}$ and $\sigma_{\varepsilon,2}$. Parameters can be estimated by sequential conditional least squares. For a fixed threshold $\gamma$ the observations may be divided into two samples $\{y_t | y_{t-d} \leq \gamma\}$ and $\{y_t | y_{t-d} > \gamma\}$: the data can be denoted respectively as $\mathbf{y}_j = (y_{ji_1}, y_{ji_2}, ..., y_{ji_{N_j}})'$ in regimes $j = 1, 2$, with $N_1$ and $N_2$ be the regimes sample sizes and $N_1 + N_2 = N - \max(p, d)$.

Parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ can be estimated by OLS as

$$\hat{\boldsymbol{\theta}}_j = \left(\mathbf{X}_j^{*\prime}\mathbf{X}_j^*\right)^{-1} \mathbf{X}_j^{*\prime}\mathbf{y}_j \tag{2}$$

for $j = 1, 2$ where $\mathbf{X}_j^* = (\mathbf{X}_j, \mathbf{Z}_j) = ((\mathbf{x}_{ji_1}', ..., \mathbf{x}_{ji_{N_j}}')', (\mathbf{z}_{ji_1}', ..., \mathbf{z}_{ji_{N_j}}')')$ is the $(N_j \times (p+r))$ matrix of regressors for each regime. The variance estimates can be calculated as $\hat{\sigma}_{\varepsilon,j} = \mathbf{r}_j'\mathbf{r}_j/(N_j - (p+r))$, with $\mathbf{r}_j = \mathbf{y}_j - \mathbf{X}_j^*\hat{\boldsymbol{\theta}}_j$.

The least square estimate of $\gamma$ is obtained by minimizing the joint residual sum of squares

$$\gamma = \arg\min_{\gamma \in \Gamma} \sum_{j=1}^{2} \mathbf{r}_j'\mathbf{r}_j \tag{3}$$

over a set $\Gamma$ of allowable threshold values so that each regime contains at least a given fraction $\varphi$ (ranging from 0.05 to 0.3) of all observations[2].

*2.2. Robust estimation of SETAR models*

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:

$$\hat{\boldsymbol{\theta}}_j^{(n+1)} = \left(\mathbf{X}_j^{*\prime}\mathbf{W}_j^{(n)}\mathbf{X}_j^{*}\right)^{-1}\mathbf{X}_j^{*\prime}\mathbf{W}_j^{(n)}\mathbf{y}_j \tag{4}$$

where $\hat{\boldsymbol{\theta}}_j^{(n+1)}$ is the GM estimate for the parameter vector in regime $j = 1, 2$ after the $n$-th iteration from an initial estimate $\hat{\boldsymbol{\theta}}_j^{(0)}$, and $\mathbf{W}_j^{(n)}$ is a weight diagonal $(N_j \times N_j)$ matrix, whose elements depend on a weighting function $w(\hat{\boldsymbol{\theta}}_j^{(n)}, \hat{\sigma}_{\varepsilon,j}^{(n)})$ bounded between 0 and 1. The threshold $\gamma$ can be estimated by minimizing the objective function $\rho(\mathbf{r}_1, \mathbf{r}_2)$ over the set $\Gamma$ of allowable threshold values.

Different weight functions have been proposed in the literature. The first method is described in Chan & Cheung (1994). Weights are calculated as

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

where $m_{y,j}$ 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. In this case, $\psi$ is the redescending Tukey bisquare weight function, defined as

$$\psi(u) = \begin{cases} (1-(u/c)^2)^2 & \text{if } |u| \leq c, \\ 0 & \text{if } |u| > c. \end{cases}$$

where $c$ is the tuning constant taken equal to 1 following Chan & Cheung (1994). The objective function to minimize for the search of the threshold depends on Tukey bisquare weights. We use the same function described in Chan & Cheung (1994).

[2] In order to ensures a sufficient number of observations around the true threshold parameter so that it can be identified, the value of $\varphi$ is usually set between 0.10 and 0.15 (Gonzalo & Pitarakis, 2002). In the simulation study of section 3 and in the applied study of section 4 we have used a value of $\varphi = 0.15$ which make the OLS estimation of the threshold "naturally" robust and more difficult to outperform by the robust estimators. Moreover, 0.15 is the default value used by the `selectSETAR` R function of the library `tsDyn`.

For the second method, we follow Franses & van Dijk (2000). The GM weights are presented in Schweppe's form $w(\hat{\boldsymbol{\theta}}_j, \hat{\sigma}_{\varepsilon,j}) = \psi(r_t)/r_t$ with standardized residuals $r_t = (y_t - \mathbf{x}_t^* \hat{\boldsymbol{\theta}}_j)/(\hat{\sigma}_{\varepsilon,j} w(\mathbf{x}_t^*))$ and $w(\mathbf{x}_t^*) = \psi(\mathrm{d}(\mathbf{x}_t^*)^\alpha)/\mathrm{d}(\mathbf{x}_t^*)^\alpha$. $\mathrm{d}(\mathbf{x}_t^*) = |\mathbf{x}_t^* - m_{y,j}|/\hat{\sigma}_{y,j}$ is the Mahalanobis distance and $\alpha$ is a constant usually set equal to 2 to obtain robustness of standard errors. The chosen weight function is the Polynomial $\psi$ function as proposed in Lucas et al. (1996), given by

$$\psi(u) = \begin{cases} u & \text{if} \quad |u| \leq c_1, \\ \mathrm{sgn}(u) g(|u|) & \text{if} \quad c_1 < |u| \leq c_2, \\ 0 & \text{if} \quad |u| > c_2, \end{cases}$$

where $\mathrm{sgn}(u)$ is the signum function, $g(|u|)$ is a fifth-order polynomial such that $\psi(u)$ is twice continuously differentiable, and $c_1$ and $c_2$ are tuning constants, taken to be the square roots of the 0.99 and 0.999 quantiles of the $\chi^2(1)$ distribution ($c_1 = 2.576$ and $c_2 = 3.291$)[3]. The threshold $\gamma$ is estimated by minimizing the objective function $\sum_{t=1}^{N} w(\hat{\boldsymbol{\theta}}, \hat{\sigma}_\varepsilon)(y_t - \mathbf{x}_t^* \hat{\boldsymbol{\theta}})^2$ over the set $\Gamma$ of allowable threshold values.

The third method is based on the same methodologies of the second but with $\psi$ be the Huber weight function, given by

$$\psi(u) = \begin{cases} -c & \text{if} \quad u \leq -c, \\ u & \text{if} \quad -c < u \leq c, \\ c & \text{if} \quad u > c, \end{cases}$$

where $c$ is a tuning constant taken equal to 1.345 to produce an estimator that has an efficiency of 95 per cent compared to the OLS estimator if $\varepsilon_t$ is normally distributed.

## 3. Simulation experiment

In their original paper Chan & Cheung (1994) carried out a simulation study to evaluate the bias of OLS and GM estimators of SETAR parameters. The simulation experiment was based on quite short time series ($N = 100$) generated from eighteen different SETAR processes. The outliers were included considering a simple pattern based on few values of the contamination parameter.

[3] Different values of the tunig constants have been used but results both of simulations and forecasting does not seem to be strongly influenced.

Finally, they considered just the Tukey's weighting function without any comparison with other possible weighting functions. The Monte Carlo simulation performed in this paper extends the Chan & Cheung (1994)'s experiment in three directions:

- two additional sample size are considered, that is $N = 500$ and $N = 1000$;
- more complex contamination patterns are analyzed: one single outlier and three outliers for all sample sizes, multiple outliers at fixed positions and at random positions for large sample sizes ($N = 500$ and $N = 1000$).
- two new weighting functions (the Huber's and the polynomial function) are applied to obtain new robust GM estimators whose performances are compared to those of the Tukey's function.

To assess the performance of the three weighting functions, we reproduce the simulation study of Chan & Cheung (1994) using the same eighteen combinations of parameters $\boldsymbol{\theta} = (\beta_1, \beta_2, \gamma, d)$ to simulate from the same processes used by Chan & Cheung (1994)[4]. We generate time series from SETAR(1,$d$) models for fixed sample sizes of $N = 100, 500, 1000$, with 1000 replications respectively, and $\sigma_\varepsilon^2 = 1$.

The series are contaminated following four schemes. For the single-outlier case, applied only for series with $N = 100$, an additive outlier is located at $t = N/2$ with magnitude $\omega = 0, 3, 4, 5$ times the standard deviation of the process. For the 3-outlier case ($N = 100, 500$), we fixed three outliers at $t = N/4, N/2$, and $N * 3/4$ with magnitude $-\omega, \omega, -\omega$ respectively. The multiple-outlier case is applied only for series with $N = 500$: three outliers are fixed every 100 observations with the same scheme of the 3-outlier case. The fourth scheme is reserved to series with a sample size of $N = 1000$: a random outlier contamination obtained using a binomial distribution with the fixed probability of 4%.

For the first robust estimation method based on the Tukey's weighting function, following Chan & Cheung (1994), the starting values $\beta_1^0, \beta_2^0$ of the parameters are calculated by four iterations with Huber weights with OLS estimates as initial points. For the second and third method based on the polynomial and the Huber's function, respectively, the starting values are calculated by least median squares[5].

[4]See Chan & Cheung (1994) for the 18 parameter combinations.

[5]Different starting values have been chosen deliberately to keep the first method as it was originally suggested by Chan & Cheung (1994).

Table 1: *Number of cases (out of 18) RMSEs of the Robust estimation are better than RMSEs of the LS estimation. 1000 MC simulations of time series with sample sizes* $N = 100, 500, 1000$ *and different contamination patterns. First column reports the name of the weighting function.*

| | $\hat{\gamma}$ | | | | $\hat{\beta}_1$ | | | | $\hat{\beta}_2$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Case | $\omega = 0$ | 3 | 4 | 5 | $\omega = 0$ | 3 | 4 | 5 | $\omega = 0$ | 3 | 4 | 5 |
| **Sample size** $N = 100$ | | | | | | | | | | | | |
| **Single-outlier case** | | | | | | | | | | | | |
| POL | 3 | 3 | 5 | 6 | 0 | 0 | 5 | 13 | 0 | 10 | 13 | 13 |
| HUB | 4 | 4 | 4 | 7 | 0 | 5 | 12 | 15 | 0 | 11 | 14 | 14 |
| TUK | 2 | 2 | 2 | 4 | 0 | 1 | 3 | 4 | 0 | 2 | 7 | 10 |
| **3-outlier case** | | | | | | | | | | | | |
| POL | 2 | 4 | 6 | 6 | 0 | 9 | 14 | 15 | 0 | 12 | 14 | 14 |
| HUB | 3 | 4 | 6 | 6 | 0 | 12 | 15 | 16 | 0 | 13 | 14 | 14 |
| TUK | 2 | 2 | 2 | 2 | 0 | 11 | 11 | 11 | 0 | 3 | 11 | 11 |
| **Sample size** $N = 500$ | | | | | | | | | | | | |
| **3-outlier case** | | | | | | | | | | | | |
| POL | 4 | 2 | 6 | 5 | 0 | 11 | 14 | 17 | 0 | 12 | 14 | 14 |
| HUB | 4 | 3 | 4 | 6 | 0 | 12 | 16 | 17 | 0 | 13 | 14 | 15 |
| TUK | 0 | 0 | 1 | 1 | 0 | 1 | 2 | 4 | 0 | 0 | 1 | 1 |
| **Multiple-outlier case** | | | | | | | | | | | | |
| POL | 4 | 4 | 8 | 10 | 0 | 15 | 17 | 18 | 0 | 14 | 15 | 16 |
| HUB | 4 | 5 | 7 | 7 | 0 | 17 | 18 | 18 | 0 | 14 | 16 | 16 |
| TUK | 0 | 2 | 2 | 2 | 0 | 9 | 13 | 13 | 0 | 11 | 14 | 14 |
| **Sample size** $N = 1000$ | | | | | | | | | | | | |
| **Random outliers contamination** (4%) | | | | | | | | | | | | |
| POL | 4 | 4 | 7 | 9 | 0 | 18 | 18 | 18 | 0 | 17 | 18 | 18 |
| HUB | 4 | 7 | 7 | 8 | 0 | 18 | 18 | 18 | 0 | 17 | 17 | 17 |
| TUK | 0 | 2 | 2 | 2 | 0 | 12 | 13 | 13 | 0 | 13 | 12 | 13 |

In Table 1 we have summarized the results of the Monte Carlo experiment. The purpose of this Table is to examine how many times each of the three robust GM estimators, called "TUK" (Tukey), "POL" (Polynomial) and "HUB" (Huber), give better estimation results of the non-robust LS estimator in terms of Root Mean Squared Error (RMSE). Three parameters (the threshold $\gamma$ and the two AR parameters $\beta_1$ and $\beta_2$) are estimated on trajectories generated without contamination and with different levels of contamination ($\omega = 0, 3, 4, 5$).

The main results can be summarized as follows. When the series are not contaminated ($\omega = 0$), LS is expected to better estimate the parameters. For this reason, the RMSE of the autoregressive parameters estimated by the robust estimators is never lower than the RMSE of the LS estimator. As regards the threshold parameter, only few times the RMSE of the HUB and POL is smaller than that of the LS. According to what it has been proven by Zhang et al. (2009), the robust estimators of the threshold parameter are less efficient than the LS estimator in small samples. As a consequence, we found that all three robust methods performed generally worse than the LS, at least for weak contamination patterns, that is in the single outlier case with small magnitude ($\omega = 3$).

Increasing the sample size and the complexity of the contamination pattern, the robust estimation of the autoregressive parameters becomes increasingly better than the LS method. For instance, moving form $N = 100$ to $N = 500$ the number of times when HUB and POL estimate the autoregressive parameters better than LS varies between 14 and 17 out of 18 with a 3-outlier contamination and $\omega \geq 4$. The number of success reach the maximum value (18) when $N = 1000$ and 4% contamination is introduced (lower panel of Table 1). The same results are not shown by the TUK's estimator, whose performances are always lower than HUB and POL and many times are even worse than those of the LS estimator.

Drawing our attention on the threshold parameter ($\gamma$, first columns of Table 1), it is immediately clear that, while the method suggested by Chan & Cheung (1994) based on the Tukey function does not show any significant improvement with respect to LS, the other two methods look to be competitive to LS, particularly for large sample sizes and complex contamination patterns. The robust estimation of the threshold looks to be a critical issue. However, we need to remember that this parameters is intrinsically robust, even when the LS estimator is applied, because it is estimated on the central part of the distribution, after the removal of possible extreme observation in the queues of the distribution (see equation 3). Moreover, a more reliable comparison between the different estimators should quantify, not only the number of times a method is better than the other, but also the relative value of the RMSE. Such a comparison is shown in Table 2.

To give an overall idea of the results reported in Table 2, we have computed the average values of the RMSEs ratios of the robust estimators with respect to the LS estimator using all 18 simulated processes with 1000 MC simulations each with sample sizes $N = 100, 500, 1000$ and different contamination designs. For instance, the first value in Table 2 (1.301) means that the average value of the RMSE obtained on the 18 simulated processes with sample size $N = 100$ using the Polynomial weight function is 30.1% higher than the RMSE of the LS estimator when the threshold is estimated on non-contaminated trajectories in accordance to the higher efficiency of LS. Thus, values greater than 1 mean that the analyzed estimator is worse than the compared estimator. From Table 2 we can conclude that all robust estimators are overperformed by the LS estimator when the parameters are estimated on non-contaminated series ($\omega = 0$). However, the Polynomial function is the only one to overperform the LS estimator in the estimation of the threshold parameter when the magnitude of the contamination is high ($\omega \geq 4$) and/or the number of outliers is high. On the other hand, POL and HUB functions are always far better than LS in the estimation of $\beta_i, i = 1, 2$ on contaminated series. These results confirm the theoretical results provided by Zhang et al. (2009).

Once it has been shown that robust GM-estimators perform better than LS when long series are not-trivially contaminated, we need to choose which weighting function gives the most reliable estimates. To this purpose we compare the couples of weighting functions that could be created from the three considered in the present paper. Results are shown in Table 3 and Table 4. The clear preference of Polynomial and Huber functions to the Tukey weights is strongly confirmed. Moreover, Polynomial reveals to be always better than Huber function when the sample size increases and the magnitude and/or the number of outliers are high. In the other cases the two weighting functions look to perform quite similar. However, when the sample size is $\geq 500$ and the contamination pattern is complex (multiple-outlier case and random outlier contamination), the Polynomial function is better than Huber's function. In particular, looking at the bottom lines of Table 4, we can note that the ratio of the Polynomial RMSE to the Huber RMSE is always less than one, thus the Polynomial weighting function reveals to be the best robust estimator. In

Table 2: *Means of the 18 RMSEs ratios of the GM estimate to the LS estimate. 1000 MC simulations of time series with sample sizes* $N = 100, 500, 1000$ *and different contamination designs. First column reports the name of the weight function.*

| | $\hat{\gamma}$ | | | | $\hat{\beta}_1$ | | | | $\hat{\beta}_2$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Case | $\omega = 0$ | 3 | 4 | 5 | $\omega = 0$ | 3 | 4 | 5 | $\omega = 0$ | 3 | 4 | 5 |
| **Sample size** $N = 100$ | | | | | | | | | | | | |
| **Single-outlier case** | | | | | | | | | | | | |
| POL | 1.301 | 1.252 | 1.173 | 1.121 | 1.321 | 1.176 | 1.088 | 0.998 | 1.381 | 1.204 | 1.038 | 0.951 |
| HUB | 1.229 | 1.174 | 1.155 | 1.096 | 1.191 | 1.079 | 1.025 | 0.935 | 1.252 | 1.077 | 0.963 | 0.866 |
| TUK | 1.753 | 1.65 | 1.588 | 1.48 | 1.648 | 1.488 | 1.394 | 1.242 | 1.733 | 1.437 | 1.305 | 1.201 |
| **3-outlier case** | | | | | | | | | | | | |
| POL | 1.292 | 1.171 | 1.12 | 1.086 | 1.308 | 0.982 | 0.817 | 0.721 | 1.365 | 1.054 | 0.885 | 0.836 |
| HUB | 1.218 | 1.139 | 1.124 | 1.126 | 1.194 | 0.88 | 0.759 | 0.674 | 1.238 | 0.977 | 0.879 | 0.802 |
| TUK | 1.742 | 1.543 | 1.498 | 1.435 | 1.656 | 1.125 | 0.997 | 0.901 | 1.724 | 1.302 | 1.157 | 1.09 |
| **Sample size** $N = 500$ | | | | | | | | | | | | |
| **3-outlier case** | | | | | | | | | | | | |
| POL | 1.385 | 1.226 | 1.135 | 1.07 | 1.164 | 0.94 | 0.779 | 0.642 | 1.255 | 1.064 | 0.897 | 0.75 |
| HUB | 1.265 | 1.179 | 1.139 | 1.067 | 1.112 | 0.912 | 0.756 | 0.632 | 1.173 | 1.014 | 0.863 | 0.723 |
| TUK | 4.048 | 3.586 | 3.186 | 2.841 | 3.086 | 2.341 | 1.972 | 1.623 | 3.068 | 2.599 | 2.221 | 1.908 |
| **Multiple-outlier case** | | | | | | | | | | | | |
| POL | 1.371 | 1.088 | 0.948 | 0.885 | 1.158 | 0.612 | 0.42 | 0.33 | 1.253 | 0.67 | 0.481 | 0.392 |
| HUB | 1.278 | 1.073 | 1.02 | 1.007 | 1.115 | 0.611 | 0.451 | 0.366 | 1.173 | 0.667 | 0.513 | 0.444 |
| TUK | 4.152 | 2.939 | 2.384 | 2.079 | 3.085 | 1.109 | 0.869 | 0.765 | 3.121 | 1.485 | 1.202 | 1.057 |
| **Sample size** $N = 1000$ | | | | | | | | | | | | |
| **Random outliers contamination** (4%) | | | | | | | | | | | | |
| POL | 1.34 | 0.955 | 0.873 | 0.827 | 1.128 | 0.404 | 0.29 | 0.237 | 1.176 | 0.365 | 0.278 | 0.231 |
| HUB | 1.286 | 1.043 | 1.032 | 1.012 | 1.107 | 0.427 | 0.325 | 0.276 | 1.131 | 0.401 | 0.314 | 0.272 |
| TUK | 6.699 | 4.525 | 3.711 | 2.864 | 4.19 | 1.162 | 0.968 | 0.888 | 4.088 | 1.401 | 1.266 | 1.098 |

order to assess the performance of the Polynomial function compared to the LS estimator even in presence of strongly contaminated trajectories, Appendix A contains some tables reporting the ratio of the RMSE of the two estimators (Polynomial is the numerator) in the three-outlier case (Table A.1) and the multiple-outlier case (Table A.2). Differently from previous tables, detailed output for each generated process is reported. In most of the cases the ratio is lower than 1, so that the superiority of the polynomial on the LS estimator is confirmed. A summary of the two tables is shown in Table A.3.

As it will be discussed in section 4, series of electricity prices are usually longer than 500 times and the presence of spikes usually reproduce the most complex contamination patterns described in the present section, thus the robust GM-estimator based on the Polynomial weighting function will be used in the application.

## 4. Robust price forecasting on the Italian electricity market

### *4.1. Data description*

Following the results of the simulation experiment, in this section, we apply LS and the robust POL weighting functions to estimate parameters of SETAR models on the Italian electricity price data (*PUN, prezzo unico nazionale*), downloaded from the website of the Italian electricity authority [6]. Moreover, a comparison of the prediction accuracy of the two estimators is implemented.

The time series of prices used in the present work covers the period from January 1st, 2013 to December 31th, 2015 (26,280 data points, for $N$ = 1,095 days): year 2015 has been left for out-of-sample forecasting. The data have an hourly frequency, therefore each day consist of 24 load periods with 00:00–01:00am defined as period 1. Spot price is denoted as $P_{th}$, where $t$ specifies the day and $j$ the load period ($t = 1, 2, ..., N$; $h = 1, 2, ..., 24$).

In this study, following a widespread practice in literature (Weron, 2014), each hourly time series is modeled separately. There are at least two motivations behind this choice. First, electricity prices are generated through a day-ahead auction mechanism where equilibrium prices are obtained for each hour of next day. As different bids for each hour of next day are unknown when the auction takes place, it is then sensible to expect a stronger relation between prices observed at each hour of subsequent days, rather than between prices observed at different hours of the same day. Second,

[6] Gestore del Mercato Elettrico (GME), `http://www.mercatoelettrico.org/en/`

Table 3: *Number of cases (out of 18) RMSEs of the first robust method are better than RMSEs of the second method. 1000 MC simulations of time series with sample sizes* $N = 100, 500, 1000$ *and different contamination designs.*

| | $\hat{\gamma}$ | | | | $\hat{\beta}_1$ | | | | $\hat{\beta}_2$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Case | $\omega = 0$ | 3 | 4 | 5 | $\omega = 0$ | 3 | 4 | 5 | $\omega = 0$ | 3 | 4 | 5 |
| **Sample size** $N = 100$ | | | | | | | | | | | | |
| **Single-outlier case** | | | | | | | | | | | | |
| POL to HUB | 5 | 4 | 7 | 7 | 0 | 3 | 3 | 2 | 2 | 2 | 5 | 1 |
| POL to TUK | 18 | 16 | 17 | 18 | 14 | 14 | 14 | 14 | 16 | 16 | 17 | 17 |
| HUB to TUK | 17 | 17 | 18 | 18 | 17 | 18 | 16 | 16 | 18 | 18 | 18 | 18 |
| **3-outlier case** | | | | | | | | | | | | |
| POL to HUB | 5 | 6 | 7 | 15 | 0 | 4 | 6 | 7 | 0 | 2 | 7 | 8 |
| POL to TUK | 18 | 16 | 17 | 18 | 15 | 14 | 17 | 17 | 16 | 16 | 17 | 18 |
| HUB to TUK | 18 | 18 | 17 | 18 | 18 | 16 | 18 | 18 | 17 | 17 | 18 | 18 |
| **Sample size** $N = 500$ | | | | | | | | | | | | |
| **3-outlier case** | | | | | | | | | | | | |
| POL to HUB | 4 | 8 | 11 | 12 | 3 | 6 | 8 | 8 | 2 | 7 | 5 | 8 |
| POL to TUK | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 |
| HUB to TUK | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 |
| **Multiple-outlier case** | | | | | | | | | | | | |
| POL to HUB | 5 | 4 | 13 | 17 | 3 | 11 | 15 | 16 | 3 | 11 | 13 | 17 |
| POL to TUK | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 |
| HUB to TUK | 18 | 18 | 18 | 18 | 18 | 17 | 18 | 18 | 18 | 18 | 18 | 18 |
| **Sample size** $N = 1000$ | | | | | | | | | | | | |
| **Random outliers contamination** (4%) | | | | | | | | | | | | |
| POL to HUB | 12 | 15 | 18 | 18 | 3 | 16 | 17 | 18 | 1 | 17 | 18 | 18 |
| POL to TUK | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 |
| HUB to TUK | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 | 18 |

Table 4: *Means of the 18 RMSEs ratios of the GM estimation. 1000 MC simulations of time series with sample sizes* $N = 100, 500, 1000$ *and different contamination designs.*

| | $\hat{\gamma}$ | | | | $\hat{\beta}_1$ | | | | $\hat{\beta}_2$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Case | $\omega = 0$ | 3 | 4 | 5 | $\omega = 0$ | 3 | 4 | 5 | $\omega = 0$ | 3 | 4 | 5 |
| **Sample size** $N = 100$ | | | | | | | | | | | | |
| **Single-outlier case** | | | | | | | | | | | | |
| POL to HUB | 1.057 | 1.045 | 1.022 | 1.028 | 1.106 | 1.087 | 1.068 | 1.051 | 1.103 | 1.104 | 1.073 | 1.071 |
| POL to TUK | 0.795 | 0.792 | 0.789 | 0.793 | 0.825 | 0.813 | 0.815 | 0.817 | 0.804 | 0.824 | 0.778 | 0.744 |
| HUB to TUK | 0.764 | 0.764 | 0.778 | 0.777 | 0.748 | 0.747 | 0.763 | 0.776 | 0.728 | 0.744 | 0.722 | 0.692 |
| **3-outlier case** | | | | | | | | | | | | |
| POL to HUB | 1.047 | 1.025 | 1.003 | 0.971 | 1.101 | 1.08 | 1.032 | 1.043 | 1.104 | 1.072 | 1.007 | 1.02 |
| POL to TUK | 0.793 | 0.795 | 0.784 | 0.777 | 0.816 | 0.858 | 0.766 | 0.728 | 0.811 | 0.823 | 0.75 | 0.723 |
| HUB to TUK | 0.766 | 0.779 | 0.78 | 0.798 | 0.74 | 0.789 | 0.736 | 0.692 | 0.735 | 0.766 | 0.742 | 0.704 |
| **Sample size** $N = 500$ | | | | | | | | | | | | |
| **3-outlier case** | | | | | | | | | | | | |
| POL to HUB | 1.07 | 1.033 | 1.009 | 1.015 | 1.039 | 1.033 | 1.021 | 1.013 | 1.069 | 1.037 | 1.025 | 1.019 |
| POL to TUK | 0.52 | 0.512 | 0.51 | 0.509 | 0.436 | 0.472 | 0.462 | 0.446 | 0.442 | 0.449 | 0.436 | 0.43 |
| HUB to TUK | 0.504 | 0.509 | 0.517 | 0.52 | 0.423 | 0.459 | 0.451 | 0.437 | 0.412 | 0.436 | 0.425 | 0.426 |
| **Multiple-outlier case** | | | | | | | | | | | | |
| POL to HUB | 1.067 | 1.013 | 0.925 | 0.887 | 1.038 | 0.966 | 0.909 | 0.902 | 1.06 | 0.962 | 0.889 | 0.874 |
| POL to TUK | 0.519 | 0.504 | 0.489 | 0.491 | 0.439 | 0.586 | 0.477 | 0.419 | 0.437 | 0.52 | 0.433 | 0.399 |
| HUB to TUK | 0.507 | 0.501 | 0.524 | 0.551 | 0.424 | 0.607 | 0.522 | 0.466 | 0.414 | 0.554 | 0.491 | 0.455 |
| **Sample size** $N = 1000$ | | | | | | | | | | | | |
| **Random outliers contamination** (4%) | | | | | | | | | | | | |
| POL to HUB | 1.034 | 0.928 | 0.847 | 0.817 | 1.018 | 0.94 | 0.92 | 0.888 | 1.037 | 0.905 | 0.899 | 0.886 |
| POL to TUK | 0.428 | 0.351 | 0.352 | 0.375 | 0.336 | 0.405 | 0.329 | 0.266 | 0.324 | 0.324 | 0.241 | 0.21 |
| HUB to TUK | 0.433 | 0.386 | 0.41 | 0.453 | 0.329 | 0.424 | 0.363 | 0.307 | 0.313 | 0.355 | 0.266 | 0.239 |

it has been proven that the forecasting performances of models built on hourly prices are better than those of models estimated on average daily prices (Raviv et al., 2015).

*4.2. Preliminary adjustments and tests*

Differences in load periods can cause significant variations in price time series. A first inspection, based on graphs, spectra and ACFs (see an example in Figure 1) for different hours, shows that the series have long-run behavior 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.

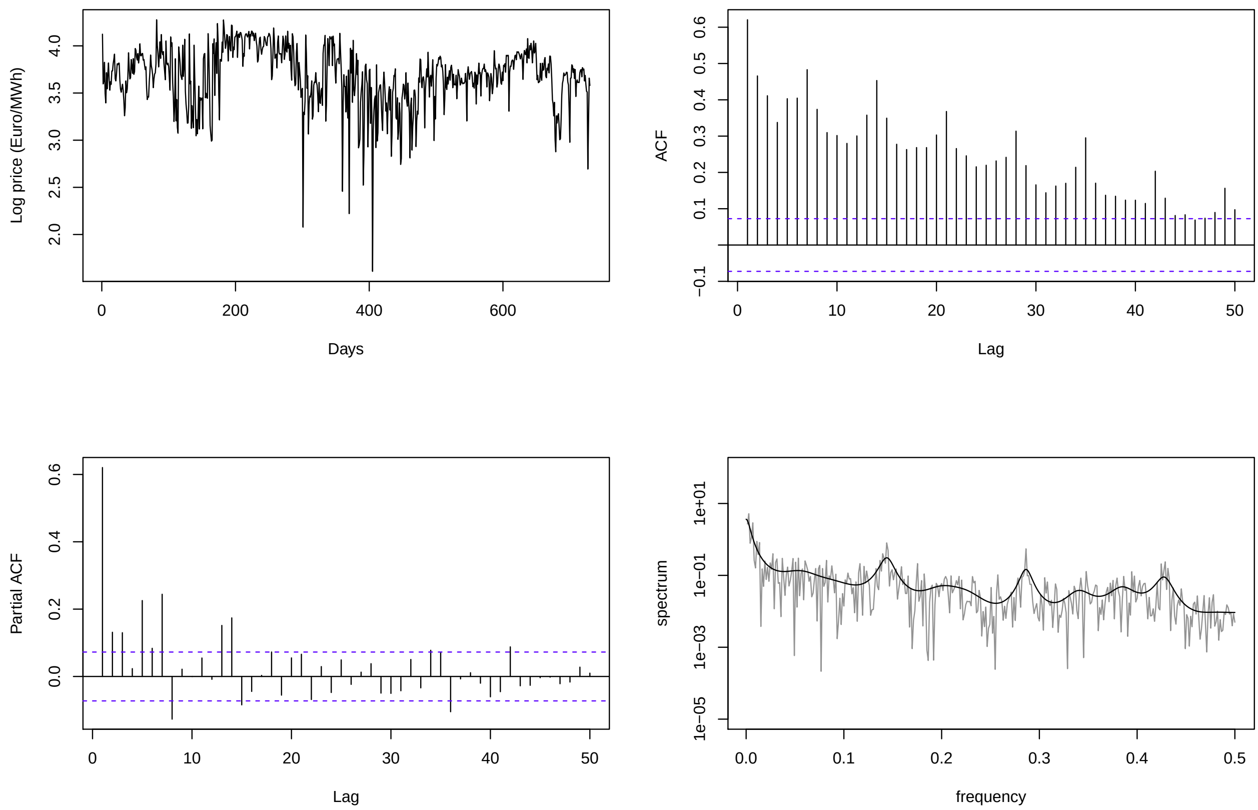


Figure 1: Time series of electricity prices on the Italian market (hour 4) from 1/1/2013 to 12/31/2014. Autocorrelations functions (ACF and PACF) and periodogram are reported.

We assume that the dynamics of log prices can be represented by a nonstationary level component $L_{th}$, accounting for level changes and/or long-term behavior, and a residual stationary component $p_{th}$, formally, $\log P_{th} = L_{th} + p_{th}$.

To estimate $L_{th}$ we used the wavelets approach (Percival & Walden, 2000). Wavelets have been used in many studies, including Trueck et al. (2007), Janczura & Weron (2010) and Lisi & Nan (2014). We considered 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 & Walden (2000)). The influence of positive and negative peaks on the estimation of $L_{th}$, has been minimized through an iterative procedure similar to that used by Nan et al. (2014) which ensures the robustness of the long-term estimation to the presence of spikes.

As an example of the time series of prices and corresponding estimated long-term component, Figure 2 shows $P_{th}$ for four different hours, with the estimated nonstationary level component superimposed[7].

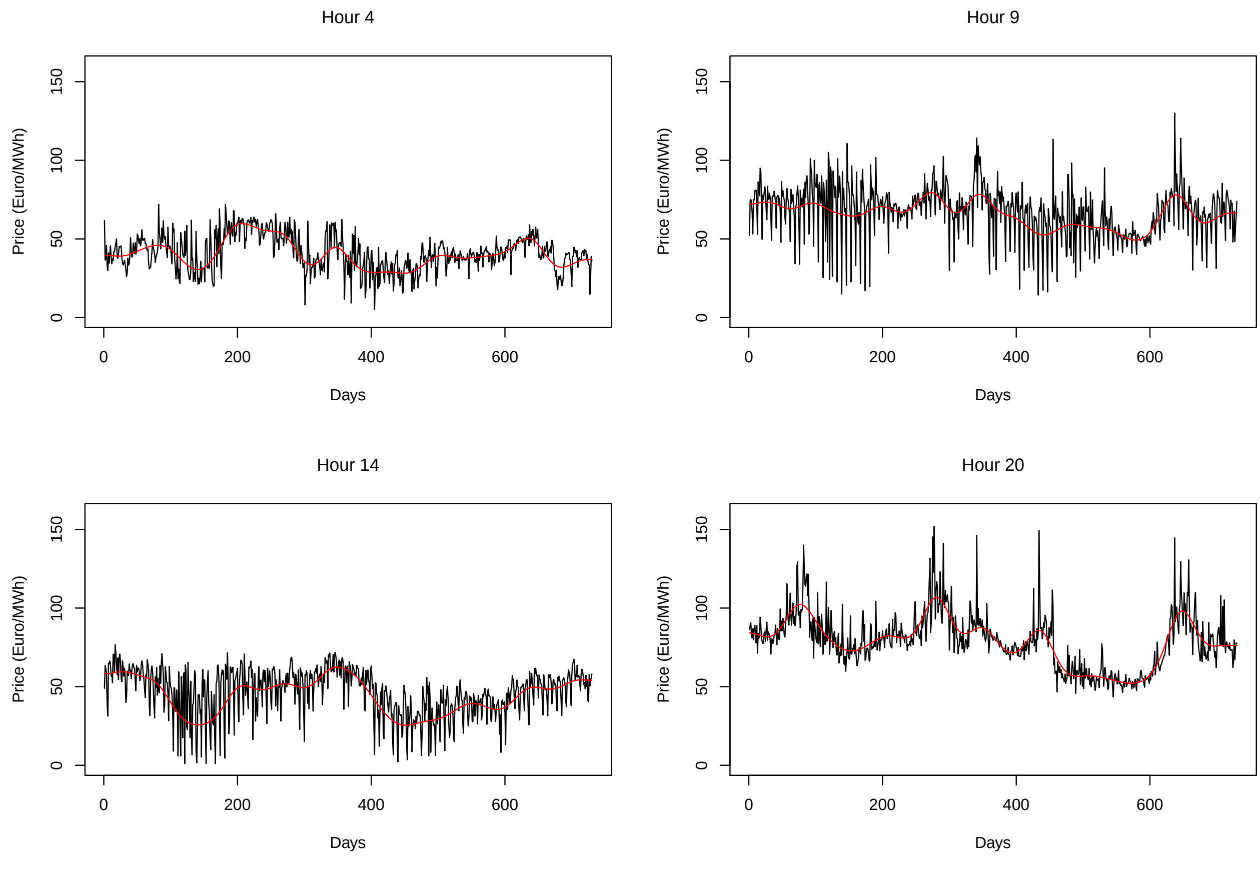


Figure 2: Long-run component (red line) estimated for four hours selected out of the total 24 hours of the sample.

[7]The remaining hours have not been reported for lack of space, but are available upon request.

It is interesting to note the different volatility structure of the time series and how the presence and magnitude of jumps changes among hours.

The time series obtained after the removal of the long-term component are stationary as it is confirmed by the application of robust and non-robust tests of unit root and stationarity. Table 5 reports the results of the application of three non-robust unit root tests, one non-robust stationarity test and one robust stationarity test. The non-robust unit-root tests are the augmented version of the Dickey-Fuller test (Said & Dickey, 1984), the Phillips-Perron test (Phillips & Perron, 1988) and the tests proposed by Elliott et al. (1996) using both the DF-GLS and the P statistics (ERS-DF-GLS and ERS-P, respectively). The stationarity test KPSS is applied both in its original non-robust version (Kwiatkowski et al., 1992) and in the robust version, recently introduced by Pelagatti & Sen (2013). The robust version of the test, based on ranks, has been computed using an auxiliary regression with 7 and 14 lags to take into account of the weekly seasonality of the data. From the Table is possible to see that, using non-robust versions of the tests (first five lines of the table), conclusions could be controversial. For example, using the ADF test with constant, in four cases the hypothesis of a unit root is rejected, even on the original time series. According to the non-robust version of the KPSS test stationarity of the original series is not rejected in 8 cases at 5% significance level. When the robust version of the KPSS is used, results are coherent and close to what it is expected: stationarity is always rejected on the original time series and almost always accepted on the de-trended series[8].

Stationary time series obtained after the long-run behavior has been removed, are suitable for the estimation of threshold models. Of course, before moving to that step, we need to test that the nonlinear threshold process could be considered a better generation process then a simpler linear model (Misiorek et al., 2006; Chan et al., 2015). As reported by Chan & Ng (2004), nonlinearity of a time series can be confounded by the presence of outliers. For this reason we applied, besides the classical F test by Tsay (1989), the robust version by Hung et al. (2009). To enhance the discriminative power of the F test in the presence of additive outliers, the Schweppe type of generalized-M (GM) estimator is considered with the polynomial weight function. Results of linearity vs. nonlinearity tests are shown in Table 6: the Table reports the number of times

[8]The tests reported in Table 5 are computed considering only a constant in the auxiliary regression, because when a linear trend has been introduced it has revealed not significant, almost in all cases.

Table 5: *Unit root and stationarity tests applied to original (log) and de-trended time series at 5% (first two columns) and 1% (last two columns) significance levels. Null hypothesis for ADF (Augmented Dickey-Fuller), PP (Phillips-Perron) and ERS (Elliot-Rothenberg-Stock) tests: presence of a unit root. Null hypothesis for KPSS (Kwiatkowski-Phillips-Schmidt-Shin, classic and robust version) tests: stationarity.*

| | Number of rejections of the Null Hypothesis | | | |
|---|---|---|---|---|
| | Significance level: 0.05 | | Significance level: 0.01 | |
| **Type of Test** | **Original** | **De-trended** | **Original** | **De-trended** |
| ADF | 4 | 24 | 2 | 24 |
| PP | 24 | 24 | 24 | 24 |
| ERS-DF-GLS | 9 | 13 | 1 | 9 |
| ERS-P | 5 | 24 | 2 | 24 |
| KPSS | 16 | 0 | 13 | 0 |
| Robust KPSS lag7 | 24 | 0 | 24 | 0 |
| Robust KPSS lag14 | 24 | 0 | 21 | 0 |

(out of the total 24 series) the hypothesis of linear generating process is rejected using both the non robust (left panel) and the robust (right panel) version of the test. Different combinations of $p$ and $d$ have been considered, taking into account the empirical autocorrelation functions of $p_{th}$ and the multilevel seasonality which is commonly shown by electricity spot prices (Janczura et al., 2013; Nowotarski et al., 2013). When daily time series of each hourly auction are analyzed, weekly frequency is the strongest source of seasonality also highlighted by the ACFs, thus, possible values of the two parameters go from 1 to 7. When the non-robust test is used, the nonlinearity hypothesis is more likely with low values of $p$, while the number of rejection increases with $p$ when the robust test is applied. However, it is immediately clear that in the majority of the cases the linearity hypothesis is rejected and the nonlinear threshold process is likely to have generated the observed trajectories, particularly when $p$ goes to 7 and the robust test is applied.

After removing the long-term component and getting, as a result, the stationary time series $p_{th}$, we are ready to estimate a SETAR($p$,$d$) model with exogenous regressors, as reported in equation (1). The order of the model (parameter $p$ and $d$) has been selected applying two robust versions of

Table 6: *F tests under the hypothesis of linearity. Number of cases the null hypothesis is rejected out of 24. Left panel: the* non robust *test by Tsay (1989) is applied. Right panel: the* robust *test by Hung et al. (2009) is applied. d is the lag of the threshold variable, p is the AR order of the model.*

| | Tsay (1989) non-robust test | | | | | | | Hung *et al.* (2009) robust test | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $d \setminus p$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | 18 | 17 | 17 | 17 | 14 | 11 | 10 | 14 | 14 | 13 | 13 | 13 | 16 | 16 |
| 2 | 18 | 10 | 8 | 17 | 17 | 16 | 10 | 14 | 12 | 12 | 12 | 13 | 16 | 17 |
| 3 | 16 | 10 | 8 | 8 | 12 | 13 | 7 | 14 | 10 | 9 | 10 | 14 | 15 | 11 |
| 4 | 18 | 17 | 17 | 11 | 10 | 13 | 15 | 12 | 15 | 15 | 13 | 14 | 14 | 16 |
| 5 | 12 | 13 | 12 | 14 | 13 | 10 | 8 | 9 | 13 | 12 | 12 | 12 | 11 | 6 |
| 6 | 9 | 9 | 11 | 13 | 13 | 15 | 15 | 13 | 17 | 17 | 17 | 16 | 8 | 10 |
| 7 | 22 | 16 | 15 | 14 | 13 | 11 | 12 | 15 | 15 | 18 | 16 | 16 | 14 | 13 |

the Akaike Information Criteria (AIC). The first proposal is based on the formula (3.8) in Franses & van Dijk (2000) for the calculation of the AIC for a 2-regime SETAR model: in our case, the variances of the regimes are calculated from the polynomial weighted residuals obtained with the robust estimation of the SETAR model. The second robust AIC proposal is contained in the paper by Tharmaratnam & Claeskens (2013) who introduce a modified information criteria based on standardized residuals obtained from MM estimates of autoregressive and scale parameters (see equation 13 of Tharmaratnam & Claeskens, 2013 and A.1 in its appendix). This AIC has been adapted for each regime to the results of the present paper by replacing the estimates with polynomial weighted estimates. The corresponding results are reported in Table 7 where the top panel refers to our first robust AIC and the bottom panel contains the output of the second robust AIC. In order to summarize the results on the 24 hours, values have been first normalized between 0 and 1 for each hour and then averaged over the 24 hours. Looking at both panels, the minimum values are observed when the threshold is estimated on $y_{t-1}$ ($d = 1$) and the 6 AR parameters are included ($p = 6$). The second minimum value is observed when $d = 1$ and $p = 7$. As prices are collected 7 days a week, weekly seasonality is more likely to be captured with $p = 7$. For this reason, a SETAR(7,1) can be considered the best generating process.

Table 7: *Robust AIC for different combinations of parameters p (columns) and d (rows). Values in the table are normalized between 0 and 1 for each hour and then averaged over the 24 hours.*

| Robust AIC based on polynomial weighted estimates | | | | | | | |
|---|---|---|---|---|---|---|---|
| $d \setminus p$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | 0.488 | 0.378 | 0.330 | 0.338 | 0.319 | 0.208 | 0.265 |
| 2 | 0.591 | 0.548 | 0.542 | 0.510 | 0.417 | 0.359 | 0.313 |
| 3 | 0.729 | 0.737 | 0.494 | 0.497 | 0.491 | 0.366 | 0.380 |
| 4 | 0.677 | 0.782 | 0.634 | 0.518 | 0.509 | 0.360 | 0.342 |
| 5 | 0.698 | 0.793 | 0.711 | 0.655 | 0.500 | 0.390 | 0.383 |
| 6 | 0.671 | 0.774 | 0.735 | 0.700 | 0.606 | 0.415 | 0.360 |
| 7 | 0.683 | 0.767 | 0.721 | 0.694 | 0.662 | 0.547 | 0.418 |

| Robust AIC based on MM estimates | | | | | | | |
|---|---|---|---|---|---|---|---|
| $d \setminus p$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | 0.564 | 0.498 | 0.475 | 0.447 | 0.452 | 0.323 | 0.344 |
| 2 | 0.647 | 0.568 | 0.584 | 0.563 | 0.485 | 0.410 | 0.419 |
| 3 | 0.636 | 0.705 | 0.564 | 0.563 | 0.537 | 0.422 | 0.376 |
| 4 | 0.668 | 0.699 | 0.682 | 0.514 | 0.541 | 0.398 | 0.345 |
| 5 | 0.652 | 0.708 | 0.630 | 0.599 | 0.541 | 0.406 | 0.442 |
| 6 | 0.686 | 0.702 | 0.669 | 0.672 | 0.575 | 0.451 | 0.422 |
| 7 | 0.665 | 0.695 | 0.634 | 0.654 | 0.649 | 0.526 | 0.419 |

*4.3. Forecasting day-ahead prices*

In section 3 we have compared the bias of different estimators of SETAR models and the superiority of robust GM-estimator (POL and HUB) has been shown and the polynomial function has been selected as the best performer. In this section, we want to compare the forecasting performances of the polynomial to those of the LS non-robust estimator.

Starting from a simple AR(7) model, which can be thought as the benchmark model, we compare the forecasting performances of the polynomial and the LS estimator, gradually increasing the complexity of the model. Thus, the basic model contains only autoregressive components, excluding the matrix $\mathbf{Z}$ reported in equation (1).

Remembering what has been said at the beginning of this section, the period 2013-2014 has been used to estimate the first model, then a set of day-ahead predictions is obtained for year 2015 applying a rolling-window procedure (see, for instance, Gianfreda & Grossi, 2012). As it is well known, the forecasting ability of models can be influenced by yearly seasons and the presence of spikes can vary from season to season. For this reason, the comparison is done not only for the whole year but also for each single season (winter: January-March, spring: April-June, summer: July-September, autumn: October-December).

The prediction ability of different models is evaluated using two different prediction error statistics: the Mean Square Error (MSE) and the Mean Absolute Error (MAE)[9]. The comparison between pairs of models is tested by means of statistical tests. The most common tests are the Diebold and Mariano's test (D-M) (Diebold & Mariano, 1995) and the Model Confidence Set test (MCS) (Hansen et al., 2003, 2011). In this paper the 1-tailed version of the Diebold-Mariano and MCS test at 5% significance level are used, considering the MSE and MAE loss functions.

In Table 8 and 9 a simple AR(7) model is compared with a SETAR(7,1), when both LS and Polynomial (POL) estimators are applied. Table 8 reports the number of times (out of the total 24 hours) the AR outperforms the SETAR model. Table 9 shows results for the opposite case. In the last row of the tables, the fraction of cases in which one model is better than the other (out of the 120 cases[10]) is computed. Summing up the numbers of the last row in the two tables we get

[9] We didn't use the "percentage" version of MSE and MAE because in 2015 prices very close to zero was observed which could heavily bias the values of MSPE and MAPE.

[10] The total number of possible cases is given by $24 \times 5 = 120$, where 24 is the number of load periods in a day and 5 is the sum of the four seasons and the whole year.

100 for MSE and MAE, while the result is lower than 100 for the two tests (D-M and MCS test) because only significant cases are included. For instance, looking at row labeled "Whole" in Table 8, we argue that in 7 hours (load periods) the AR(7) estimated by LS performs better than the SETAR(7,1), estimated by LS, when the day-ahead forecasts for the whole year are included in the computation of MSE. Of course, the number in the same position, but in Table 9 is the complement to 24, that is 17. If we stay on the same row ("Whole") but focus on the D-M test columns, we find that just in 2 cases the forecasting performance of the AR(7) model is significantly better than the performance of the SETAR(7,1) using the MSE as loss function and the LS estimator. The number found in the same position, but in Table 9 is not the complement of 2 to 24, but 7, meaning that in 7 load periods the SETAR is significantly better than the AR model. In the remaining cases $(24 - 7 - 2 = 15)$ none of the two models significantly outperforms the other. Focusing on the last line of both tables is possible to conclude that the nonlinearity of SETAR model enables to better predict electricity prices in most of the cases, thus confirming the output of nonlinearity tests (see Table 6).

Tables 10 and 11 compare the forecasting ability of the LS and POL estimator of the basic SETAR(7,1) model, without external regressors. The superiority of the robust estimator (POL) is quite clear, particularly when all days of the year are included. In this case, in 22 cases the Predictor Error Statistics (MSE and MAE) of POL are lower than those of LS and in 14 cases the performance of POL is significantly better than that of LS applying the Diebold-Mariano test (Table 11). The preference for the robust estimator on LS is not so clear in spring (April-June period), but this is due to the low presence of spikes in that time span and confirms the higher efficiency of LS with respect to robust estimators for uncontaminated series (see section 3).

The superiority of the robust estimator is overwhelming when regressors are introduced.

In the literature on electricity price forecasting, the strong influence of exogenous regressor on model's forecasting performances has been widely discussed (Gianfreda & Grossi, 2012; Weron, 2014). For this reason, we need to draw our attention on the possibility to introduce regressors which could improve the forecasting ability of the model by catching the peculiarities of the market. With reference of the Italian market, and taking the availability of predicted exogenous regressors into account, the following set of regressors are introduced in the models:

- deterministic day-of-the-week dummy variables, that is $D_k$, with $k = 1, \ldots, 6$;

Table 8: *Number of cases AR model gives better results than SETAR model (four seasons and whole year 2015), LS and POL estimation. Comparisons with prediction error statistics (PES) values and p-values for the 1-tailed Diebold-Mariano and MCS tests at 5% significance level, MSE and MAE loss functions.*

| | **PES Ratios** | | | | **D-M test** | | | | **MCS test** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **MSE** | | **MAE** | | **MSE** | | **MAE** | | **MSE** | | **MAE** | |
| **Period** | LS | POL | LS | POL | LS | POL | LS | POL | LS | POL | LS | POL |
| Jan-Mar | 8 | 10 | 8 | 9 | 0 | 3 | 2 | 2 | 0 | 1 | 1 | 1 |
| Apr-Jun | 12 | 14 | 6 | 16 | 0 | 3 | 0 | 3 | 0 | 1 | 0 | 2 |
| Jul-Sep | 6 | 3 | 6 | 3 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| Oct-Dec | 7 | 7 | 7 | 7 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 |
| Whole | 7 | 7 | 5 | 7 | 2 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| Totals (120 cases) | 33.33% | 34.17% | 26.67% | 35.00% | 3.33% | 7.50% | 3.33% | 5.83% | 0.00% | 1.67% | 1.67% | 2.50% |

Table 9: *Number of cases SETAR model gives better results than AR model (four seasons and whole year 2015), LS and POL estimation. Comparisons with prediction error statistics (PES) values and p-values for the 1-tailed Diebold-Mariano and MCS tests at 5% significance level, MSE and MAE loss functions.*

| | **PES Ratios** | | | | **D-M test** | | | | **MCS test** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **MSE** | | **MAE** | | **MSE** | | **MAE** | | **MSE** | | **MAE** | |
| **Period** | LS | POL | LS | POL | LS | POL | LS | POL | LS | POL | LS | POL |
| Jan-Mar | 16 | 14 | 16 | 15 | 4 | 2 | 6 | 2 | 2 | 0 | 5 | 1 |
| Apr-Jun | 12 | 10 | 18 | 8 | 5 | 1 | 5 | 1 | 1 | 0 | 2 | 1 |
| Jul-Sep | 18 | 21 | 18 | 21 | 12 | 10 | 9 | 11 | 6 | 7 | 7 | 8 |
| Oct-Dec | 17 | 17 | 17 | 17 | 6 | 10 | 6 | 7 | 3 | 5 | 3 | 4 |
| Whole | 17 | 17 | 19 | 17 | 7 | 9 | 10 | 9 | 7 | 6 | 7 | 7 |
| Totals (120 cases) | 66.67% | 65.83% | 73.33% | 65.00% | 28.33% | 26.67% | 30.00% | 25.00% | 15.83% | 15.00% | 20.00% | 17.50% |

Table 10: *SETAR model: number of cases LS model gives better results than POL model (four seasons and whole year 2015). Comparisons with prediction error statistics (PES) values and p-values for the 1-tailed Diebold-Mariano and MCS tests at 5% significance level, MSE and MAE loss functions.*

| | **PES Ratios** | | **D-M test** | | **MCS test** | |
|---|---|---|---|---|---|---|
| **Period** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** |
| Jan-Mar | 12 | 5 | 1 | 0 | 0 | 0 |
| Apr-Jun | 6 | 6 | 0 | 0 | 0 | 0 |
| Jul-Sep | 4 | 5 | 0 | 0 | 0 | 0 |
| Oct-Dec | 5 | 5 | 1 | 1 | 0 | 1 |
| Whole | 2 | 2 | 0 | 0 | 0 | 0 |
| Totals (120 cases) | 24.17% | 19.17% | 1.67% | 0.83% | 0.00% | 0.83% |

Table 11: *SETAR model: number of cases POL model gives better results than LS model (four seasons and whole year 2015). Comparisons with prediction error statistics (PES) values and p-values for the 1-tailed Diebold-Mariano and MCS tests at 5% significance level, MSE and MAE loss functions.*

| | **PES Ratios** | | **D-M test** | | **MCS test** | |
|---|---|---|---|---|---|---|
| **Period** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** |
| Jan-Mar | 12 | 19 | 1 | 2 | 0 | 2 |
| Apr-Jun | 18 | 18 | 2 | 2 | 2 | 1 |
| Jul-Sep | 20 | 19 | 11 | 10 | 6 | 7 |
| Oct-Dec | 19 | 19 | 5 | 6 | 6 | 5 |
| Whole | 22 | 22 | 14 | 14 | 9 | 7 |
| Totals (120 cases) | 75.83% | 80.83% | 27.50% | 28.33% | 19.17% | 18.33% |

- day-ahead predicted demand of electricity, made available by the Italian authority (GME);
- day-ahead predicted wind generation, made available by the Italian Transmission System Operator (TSO) Terna.[11]

Tables 12 and 13 compare the predictive accuracy of LS and POL estimators for the complex model SETARX(7,1) containing the above exogenous regressors. In this model, matrix **Z** contains the detrended day-ahead predicted demand of electricity and the detrended predicted electricity generation by wind. As for the price series, the level component of the two forecasted regressors has been estimated using the wavelets approach. Comparing Table 11 to Table 13, the fraction of cases where the POL estimator significantly outperform the LS estimator moves from less than 30% to almost 50% when the D-M test on MAE is considered.

Table 12: *SETAR with forecasted demand, dummies and forecasted wind generation: number of cases LS model gives better results than POL model (four seasons and whole year 2015). Comparisons with prediction error statistics (PES) values and p-values for the 1-tailed Diebold-Mariano and MCS tests at 5% significance level, MSE and MAE loss functions.*

| | **PES Ratios** | | **D-M test** | | **MCS test** | |
|---|---|---|---|---|---|---|
| **Period** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** |
| Jan-Mar | 4 | 2 | 1 | 0 | 0 | 0 |
| Apr-Jun | 10 | 5 | 0 | 0 | 0 | 0 |
| Jul-Sep | 7 | 7 | 0 | 0 | 0 | 0 |
| Oct-Dec | 5 | 6 | 1 | 1 | 0 | 1 |
| Whole | 4 | 3 | 0 | 0 | 0 | 0 |
| Totals (120 cases) | 25.00% | 19.17% | 1.67% | 0.83% | 0.00% | 0.83% |

## 5. Conclusions

A robust approach to modelling and forecasting electricity prices is suggested. As it is well known, one of the main stylized facts observed on electricity spot markets is the presence of

[11] https://www.terna.it/en-gb/home.aspx

Table 13: *SETAR with forecasted demand, dummies and forecasted wind generation: number of cases POL model gives better results than LS model (four seasons and whole year 2015). Comparisons with prediction error statistics (PES) values and p-values for the 1-tailed Diebold-Mariano and MCS tests at* 5% *significance level, MSE and MAE loss functions.*

| | **PES Ratios** | | **D-M test** | | **MCS test** | |
|---|---|---|---|---|---|---|
| **Period** | **MSE** | **MAE** | **MSE** | **MAE** | **MSE** | **MAE** |
| Jan-Mar | 20 | 22 | 12 | 13 | 10 | 12 |
| Apr-Jun | 14 | 19 | 5 | 6 | 3 | 7 |
| Jul-Sep | 17 | 17 | 7 | 10 | 3 | 8 |
| Oct-Dec | 19 | 18 | 12 | 12 | 8 | 10 |
| Whole | 20 | 21 | 16 | 18 | 15 | 18 |
| Totals (120 cases) | 75.00% | 80.83% | 43.33% | 49.17% | 32.50% | 45.83% |

sudden departure of prices from the normal regime for a very short time interval. This particular pattern is usually called "spike". While the literature on electricity prices has so far focused on the modelling and prediction of spikes, this paper has dealt with robust estimators of models for electricity prices. Robust estimators are not strongly affected by the presence of spikes and are effective in the prediction of "normal" prices which are the majority of the data observed on electricity markets.

Another stylized fact observed on electricity markets is the nonlinear nature of the generating processes of prices. Threshold processes of time series are particular nonlinear processes which could be robustly estimated through a generalization to dependent data of GM-estimator originally developed for independent data.

Different proposals could be found in the literature, applying GM-robust estimator to SETAR based on different weighting functions. However, the different proposals have never been deeply compared to decide which function gives the smaller bias under particular conditions.

For this reason, we have carried out a Monte Carlo experiment to compare LS and GM estimators, with different weighting functions, for SETAR models: the Tukey's function, originally proposed and studied by Chan & Cheung (1994), the Huber's function, studied by Zhang et al.

(2009) and the polynomial function of Lucas et al. (1996) suggested in Giordani (2006). The main result is that the bias in the threshold parameter estimator, which has been observed in previous works, decreases when Huber's and Polynomial weighting functions are applied, when the sample size increases and for complex contamination patterns. However, when the features of the trajectories are more similar to what is observed on electricity markets, the polynomial function looks to be the best estimator.

The robust GM-estimator of SETAR processes based on the polynomial weights has been applied to forecast hourly day-ahead spot prices observed on the Italian market in the period 2013-2015. The long-run trend has been estimated using a wavelet-based procedure and the stationarity of the de-trended series has been verified through robust tests. The nonlinearity of the generating process has been robustly tested using non-robust and robust tests. Finally the order of the SETAR model has been selected by a robust version of the Akaike Information Criteria.

Using prediction error statistics (MSE and MAE) and forecasting performance tests (Diebold and Mariano test and Model Confidence Set test), the nonlinear process SETAR(7,1) has revealed more effective than a linear AR(7) in predicting prices for year 2015, confirming the output of the robust test for nonlinearity. Besides the information set given by the past observations, several exogenous variables can used to improve the forecasting performances of nonlinear models applied to electricity prices. Following recent literature (Cló et al., 2015; Ketterer, 2014), days-of-the-week dummy variables, predicted electricity demand and predicted wind power generation have been introduced as exogenous regressors in the SETAR(7,1) model on the Italian market.

The superiority of the forecasting performance of the robust on the LS estimator with exogenous regressor is overwhelming. The introduction of effective regressors, not only improve the forecasting power of the models, but the predictive ability of the robust estimator is significantly better than that of the LS estimator in more than 50% of the total cases.

It is remarkable to stress that on the Italian market very large prices are never observed and even the highest prices collected in the last years could not be strictly defined as "spikes" in the sense used in other papers (see, for instance, Haldrup et al., 2016) applied to the Nordpool market. However, the robust estimators have revealed very effective in improving the forecasting performances of the model. Moreover, the overwhelming superiority of the method for models with regressors has proven that robust estimators are particularly desirable when multivariate extreme observations happens although spikes in univariate time series are not so clear.

Future research will be devoted to the application of robust estimators to markets other than the Italian and to study the asymptotic properties of the robust polynomial estimator when larger samples are considered.

**Acknowledgements**

The authors wish to thank people attending conferences where preliminary versions of this work were presented. The 9th Meeting of the Classification and Data Analysis Group (ClaDAG 2013), in particular Anna Clara Monti, Andrea Cerioli, Marco Riani, Luigi Spezia. People attending the CFE 2013. People attending the seminar delivered by L. Grossi on February 2014 at Bicocca university (Milan), in particular Lucia Parisio, Emilio Colombo, Matteo Manera. People attending the Mannheim Energy Conference 2014 on Sustainable Energy Markets, in particular M. Waterson, S. Heim, A. Momeni, D. Schober. People attending the 11th Conference on European Energy Market 2014, in Krakow, in particular Y. Perez, J. Nowotarski and three anonymous referees for their helpful comments to the paper presented at the conference. People attending the (SIS 2014). The authors would like to thank the Department of Economics and the University of Verona for financing the research with a FAR/2013 (ex 60%) grant. They also thank the financial support of the project MIUR PRIN MISURA - Multivariate models for risk assessment. The usual disclaimer applies.

## Appendix A. Appendix

Table A.1: *Ratios of the RMSE of the GM estimate with polynomial weights to the LS estimate. 1000 MC simulations of time series with sample size 500 and outliers with magnitude* $\omega = 10$ *times the standard deviation of the processes. 3-outlier case*

| True values | | | | $\hat{\gamma}$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ |
|---|---|---|---|---|---|---|
| $\gamma$ | $\beta_1$ | $\beta_2$ | $d$ | | | |
| 0 | 0.9 | -0.1 | 1 | 1.23 | 0.136 | 1.799 |
| 0 | 0.9 | -0.77 | 1 | 1.261 | 0.145 | 0.941 |
| 0 | -0.5 | -1 | 1 | 0.61 | 0.197 | 0.176 |
| 0 | -1 | -0.5 | 1 | 0.571 | 0.108 | 0.319 |
| 0 | 0.3 | 0.8 | 1 | 1.247 | 0.428 | 0.238 |
| 0 | 0.5 | 0.8 | 1 | 1.203 | 0.233 | 0.242 |
| 0 | -0.3 | 0.8 | 1 | 0.975 | 0.749 | 0.267 |
| 0 | -0.5 | 0.8 | 1 | 0.868 | 0.425 | 0.262 |
| 0 | 0.8 | 0.3 | 1 | 1.463 | 0.165 | 0.49 |
| 0 | 0.8 | 0.5 | 1 | 1.341 | 0.159 | 0.308 |
| 0 | 0.8 | -0.3 | 1 | 1.251 | 0.185 | 0.917 |
| 0 | 0.8 | -0.5 | 1 | 1.186 | 0.186 | 0.575 |
| 0.1 | 0.3 | 0.8 | 1 | 1.098 | 0.408 | 0.245 |
| -0.1 | 0.3 | -0.8 | 1 | 0.946 | 0.581 | 0.323 |
| 0 | 0.3 | 0.8 | 2 | 0.445 | 0.563 | 0.229 |
| 0 | 0.3 | -0.8 | 2 | 0.344 | 0.348 | 0.203 |
| 0.1 | 0.3 | 0.8 | 2 | 0.496 | 0.485 | 0.239 |
| -0.1 | 0.3 | -0.8 | 2 | 0.32 | 0.357 | 0.19 |

Table A.2: *Ratios of the RMSE of the GM estimate with polynomial weights to the LS estimate. 1000 MC simulations of time series with sample size 500 and outliers with magnitude* $\omega = 10$ *times the standard deviation of the processes. Multiple-outlier case*

| True values | | | | $\hat{\gamma}$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ |
|---|---|---|---|---|---|---|
| $\gamma$ | $\beta_1$ | $\beta_2$ | $d$ | | | |
| 0 | 0.9 | -0.1 | 1 | 1.152 | 0.053 | 3.153 |
| 0 | 0.9 | -0.77 | 1 | 1.123 | 0.052 | 0.715 |
| 0 | -0.5 | -1 | 1 | 0.552 | 0.106 | 0.081 |
| 0 | -1 | -0.5 | 1 | 0.591 | 0.067 | 0.16 |
| 0 | 0.3 | 0.8 | 1 | 0.899 | 0.297 | 0.085 |
| 0 | 0.5 | 0.8 | 1 | 0.938 | 0.165 | 0.088 |
| 0 | -0.3 | 0.8 | 1 | 0.653 | 0.597 | 0.09 |
| 0 | -0.5 | 0.8 | 1 | 0.568 | 0.345 | 0.086 |
| 0 | 0.8 | 0.3 | 1 | 1.456 | 0.073 | 0.312 |
| 0 | 0.8 | 0.5 | 1 | 1.347 | 0.075 | 0.166 |
| 0 | 0.8 | -0.3 | 1 | 1.1 | 0.071 | 0.725 |
| 0 | 0.8 | -0.5 | 1 | 1.032 | 0.074 | 0.397 |
| 0.1 | 0.3 | 0.8 | 1 | 0.873 | 0.304 | 0.089 |
| -0.1 | 0.3 | -0.8 | 1 | 0.776 | 0.286 | 0.181 |
| 0 | 0.3 | 0.8 | 2 | 0.319 | 0.352 | 0.099 |
| 0 | 0.3 | -0.8 | 2 | 0.181 | 0.254 | 0.103 |
| 0.1 | 0.3 | 0.8 | 2 | 0.34 | 0.341 | 0.096 |
| -0.1 | 0.3 | -0.8 | 2 | 0.17 | 0.248 | 0.099 |

Table A.3: *Number of cases RMSEs of the GM estimation with polynomial weights are better than RMSEs of the LS estimation. 1000 MC simulations of time series with sample size 500 and outliers with magnitude* $\omega = 10$ *times the standard deviation of the processes.*

| $\hat{\gamma}$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ |
|---|---|---|
| **3-outlier case** | | |
| 9 | 18 | 17 |
| **Multiple-outlier case** | | |
| 12 | 18 | 17 |

## References


Aggarwal, S. K., Saini, L. M., & Kumar, A. (2009). Electricity price forecasting in deregulated markets: A review and evaluation. *International Journal of Electrical Power & Energy Systems*, *31*, 13 – 22.

Andersen, R. (2008). *Modern Methods for Robust Regression*. SAGE Publications, Inc.

Battaglia, F., & Orfei, L. (2005). Outlier detection and estimation in nonlinear time series. *Journal of Time Series Analysis*, *26*, 107 – 121.

Bordignon, S., Bunn, D. W., Lisi, F., & Nan, F. (2013). Combining day-ahead forecasts for british electricity prices. *Energy Economics*, *35*, 88 – 103.

Bystrom, H. N. E. (2005). Extreme value theory and extremely large electricity price changes. *International Review of Economics & Finance*, *14*, 41 – 55.

Cartea, A., & Figueroa, M. (2005). Pricing in electricity markets: A mean reverting jump diffusion model with seasonality. *Applied Mathematical Finance*, *12*, 313 – 335.

Chan, W. S., & Cheung, S. H. (1994). On robust estimation of threshold autoregressions. *Journal of Forecasting*, *13*, 37 – 49.

Chan, W.-S., Cheung, S. H., Chow, W. K., & Zhang, L.-X. (2015). A robust test for threshold-type nonlinearity in multivariate time series analysis. *Journal of Forecasting*, *34*, 441 – 454.

Chan, W.-S., & Ng, M.-W. (2004). Robustness of alternative non-linearity tests for setar models. *Journal of Forecasting*, *23*, 215 – 231.

Christensen, T., Hurn, A., & Lindsay, K. (2012). Forecasting spikes in electricity prices. *International Journal of Forecasting*, *28*, 400 – 411.

Clements, A., Fuller, J., & Hurn, S. (2013). Semi-parametric forecasting of spikes in electricity prices. *The Economic Record*, *89*, 508 – 521.

Cló, S., Cataldi, A., & Zoppoli, P. (2015). The merit-order effect in the italian power market: The impact of solar and wind generation on national wholesale electricity prices. *Energy Policy*, *77*, 79 – 88.

Conejo, A. J., Contreras, J., Espiñola, R., & Plazas, M. A. (2005). Forecasting electricity prices for a day-ahead pool-based electric energy market. *International Journal of Forecasting*, *21*, 435 – 462.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *13*, 253 – 263.

van Dijk, D. (1999). *Smooth Transition Models: Extensions and Outlier Robust Inference*. Ph.D. thesis Erasmus University Rotterdam Rotterdam, The Netherlands.

Elliott, G., Rothenberg, T. J., & Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica*, *64*, 813 – 836.

Escribano, A., Peña, J. I., & Villaplana, P. (2011). Modelling electricity prices: International evidence. *Oxford Bulletin of Economics and Statistics*, *73*, 622 – 650.

Flatley, L., Giulietti, M., Grossi, L., Trujillo-Baute, E., & Waterson, M. (2016). *Analysing the potential economic value of energy storage*. Working Papers 2016/2 Institut d'Economia de Barcelona (IEB). URL: `https://ideas.repec.org/p/ieb/wpaper/doc2016-2.html`.

Franses, P. H., & van Dijk, D. (2000). *Non-linear time series models in empirical finance*. Cambridge University Press.

Gaillard, P., Goude, Y., & Nedellec, R. (2016). Additive models and robust aggregation for GEFCom2014 probabilistic electric load and electricity price forecasting. *International Journal of Forecasting*, *32*, 1038 – 1050.

Gianfreda, A., & Grossi, L. (2012). Forecasting Italian electricity zonal prices with exogenous variables. *Energy Economics*, *34*, 2228 – 2239.

Giordani, P. (2006). A cautionary note on outlier robust estimation of threshold models. *Journal of Forecasting*, *25*, 37 – 47.

Giordani, P., Kohn, R., & van Dijk, D. (2007). A unified approach to nonlinearity, structural change, and outliers. *Journal of Econometrics*, *137*, 112 – 133.

Gonzalo, J., & Pitarakis, J.-Y. (2002). Estimation and model selection based inference in single and multiple threshold models. *Journal of Econometrics*, *110*, 319 – 352.

Grossi, L., & Nan, F. (2015). Robust estimation of regime switching models. In I. Morlini, M. Vichi, & T. Minerva (Eds.), *Advanced Statistical Models Data Analysis* chapter 14. (pp. 125 – 135). Berlin: Springer.

Haldrup, N., Knapik, O., & Proietti, T. (2016). *A generalized exponential time series regression model for electricity prices*. CREATES Research Papers Department of Economics and Business Economics, Aarhus University.

Haldrup, N., & Nielsen, M. O. (2006). A regime switching long memory model for electricity prices. *Journal of Econometrics*, *135*, 349 – 376.

Hansen, B. (1999). Testing for linearity. *Journal of Economic Surveys*, *13*, 551 – 576.

Hansen, B. E. (1997). Inference in TAR models. *Studies in Nonlinear Dynamics & Econometrics*, *2*, 1 – 14.

Hansen, P. R., Lunde, A., & Nason, J. M. (2003). Choosing the best volatility models: The model confidence set approach. *Oxford Bulletin of Economics and Statistics*, *65*, 839 – 861.

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*, 453 – 497.

Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A., & Hyndman, R. J. (2016). Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond. *International Journal of Forecasting*, *32*, 896 – 913.

Huisman, R., & Mahieu, R. (2003). Regime jumps in electricity prices. *Energy Economics*, *25*, 425 – 434.

Hung, K. C., Cheung, S. H., Chan, W.-S., & Zhang, L.-X. (2009). On a robust test for SETAR-type nonlinearity in time series analysis. *Journal of Forecasting*, *28*, 445 – 464.

Janczura, J., Trueck, S., Weron, R., & Wolff, R. C. (2013). Identifying spikes and seasonal components in electricity spot price data: A guide to robust modeling. *Energy Economics*, *38*, 96 – 110.

Janczura, J., & Weron, R. (2010). An empirical comparison of alternate regime-switching models for electricity spot prices. *Energy Economics*, *32*, 1059 – 1073.

Ketterer, J. C. (2014). The impact of wind power generation on the electricity price in germany. *Energy Economics*, *44*, 270 – 280.

Kosater, P., & Mosler, K. (2006). Can Markov regime-switching models improve power-price forecasts? Evidence from German daily power prices. *Applied Energy*, *83*, 943 – 958.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, *54*, 159 – 178.

Laouafi, A., Mordjaoui, M., Laouafi, F., & Boukelia, T. E. (2016). Daily peak electricity demand forecasting based on an adaptive hybrid two-stage methodology. *International Journal of Electrical Power & Energy Systems*, *77*, 136 – 144.

Lisi, F., & Nan, F. (2014). Component estimation for electricity prices: Procedures and comparisons. *Energy Economics*, *44*, 143 – 159.

Lucas, A., van Dijk, R., & Kloek, T. (1996). *Outlier Robust GMM Estimation of Leverage Determinants in Linear Dynamic Panel Data Models*. Discussion Paper 94-132 Tinbergen Institute.

Lucheroni, C. (2012). A hybrid SETARX model for spikes in tight electricity markets. *Operations Research and Decisions*, *22*, 13 – 49.

Maciejowska, K., Nowotarski, J., & Weron, R. (2016). Probabilistic forecasting of electricity spot prices using factor quantile regression averaging. *International Journal of Forecasting*, *32*, 957 – 965.

Maronna, R. A., Martin, R. D., & Yohai, V. J. (2006). *Robust Statistics: Theory and Methods*. Wiley, London.

Misiorek, A., Trueck, S., & Weron, R. (2006). Point and Interval Forecasting of Spot Electricity Prices: Linear vs. Non-Linear Time Series Models. *Studies in Nonlinear Dynamics & Econometrics*, *10*, 1 – 36.

Nan, F., Bordignon, S., Bunn, D. W., & Lisi, F. (2014). The forecasting accuracy of electricity price formation models. *International Journal of Energy and Statistics*, *2*, 1 – 26.

Nogales, F. J., Contreras, J., Conejo, A. J., & Espiñola, R. (2002). Forecasting next-day electricity prices by time series models. *IEEE Transactions on Power Systems*, *17*, 342 – 348.

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

Panagiotelis, A., & Smith, M. (2008). Bayesian density forecasting of intraday electricity prices using multivariate skew t distributions. *International Journal of Forecasting*, *24*, 710 – 727.

Pelagatti, M. M., & Sen, P. K. (2013). Rank tests for short memory stationarity, .

Percival, D., & Walden, A. (2000). *Wavelet Methods for Time Series Analysis*. Cambridge University Press.

Phillips, P. C. B., & Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, *75*, 335 – 346.

Raviv, E., Bouwman, K. E., & van Dijk, D. (2015). Forecasting day-ahead electricity prices: Utilizing hourly prices. *Energy Economics*, *50*, 227 – 239.

Ricky Rambharat, B., Brockwell, A. E., & Seppi, D. J. (2005). A threshold autoregressive model for wholesale electricity prices. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *54*, 287 – 299.

Said, S. E., & Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, *71*, 599 – 607.

Sapio, A., & Spagnolo, N. (2016). Price regimes in an energy island: Tacit collusion vs. cost and network explanations. *Energy Economics*, *55*, 157 – 172.

Swider, D. J., & Weber, C. (2007). Bidding under price uncertainty in multi-unit pay-as-bid procurement auctions for power systems reserve. *European Journal of Operational Research*, *181*, 1297 – 1308.

Tharmaratnam, K., & Claeskens, G. (2013). A comparison of robust versions of the AIC based on M-, S- and MM-estimators. *Statistics*, *47*, 216 – 235.

Trueck, S., Weron, R., & Wolff, R. (2007). *Outlier treatment and robust approaches for modeling electricity spot* *prices*. MPRA Paper 4711 Hugo Steinhaus Center, Wroclaw University of Technology.

Tsay, R. S. (1989). Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association*, *84*, 231 – 240.

Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, *30*, 1030 – 1081.

Zachmann, G. (2013). A stochastic fuel switching model for electricity prices. *Energy Economics*, *35*, 5 – 13. Quantitative Analysis of Energy Markets (Eds. Gianfreda, A. and Grossi, L.).

Zhang, L. X., Chan, W. S., Cheung, S. H., & Hung, K. C. (2009). A note on the consistency of a robust estimator for threshold autoregressive processes. *Statistics and Probability Letters*, *79*, 807 – 813.