Filtering outliers in time series of electricity prices

Ilaria Lucrezia Amerise

Abstract In this paper, we describe a tool for detecting outliers in electricity prices. The tool primarily consists of a filtering procedure (Whittaker smoother) that removes unpredictable effects and captures long-term smooth variations. So as to identify outliers, we compare observed prices with smoothed ones. If the difference between the two exceeds a predetermined limit, the corresponding prices are considered anomalous and candidates for an appropriate statistical treatment. The new tool is compared with another method based on local regression.

Key words: Whittacker smoother, penalized least squares, price spikes.

1 Introduction

The frequent occurrence of irregular, abrupt, extreme movements or spikes in spot prices is one of the most salient characteristics of electricity markets. These spikes are of very short duration so that only a few traces of their influence is revealed at adjacent time points. It should be recognized that price peaks and/or troughs are natural characteristics of the electricity market and are very important for energy market participants. On the other hand, we do not expect that the choice of forecasting method to be conditioned by any mere irregularity in prices. The main purpose of the present study is to identify outlying spikes by using a non-parametric filter that returns a smoothed time series, which acts as benchmark against which moderate and extreme spike prices can be identified. The same pre-processing can be applied to other phenomena with fast dynamics, such as meteorological time series, stock market indices and exchange rates.

Ilaria Lucrezia Amerise
Università della Calabria, Dipartimento di Economia, Statistica e Finanza - Cosenza, Italy e-mail: ilaria.amerise@unical.it

Let us suppose that a time series of electricity prices consists of a, not-better-specified, underlying long-term pattern and of a non-observable random error

$$p_t = \widehat{p}_t + a_t, \qquad t = 1, 2, \cdots, n \tag{1}$$

where p_t is the price in €/MWh (if necessary, log-transformed) on day t, n is the length of the time series, $\widehat{p}_t$ is the baseline and the errors a_t have zero mean. The baseline $\widehat{p}_t$ aims to capture any predictable variation in electricity price behavior arising from regularities over time and it is used to interpolate or filter the observed prices. The stochastic term a_t can be interpreted as what remains of the price after the baseline has been removed. Deviations between observed and baseline prices falling outside a prefixed range will indicate aberrant prices.

To keep (1) simple and general, the description of the baseline has to be smooth, with no preconceptions about its form. Perhaps the simplest smoothing function is a straight line parallel to the time axis. The smoothness, however, is just one side of the coin. The filtered prices must also be close to the observed prices. For example, a polynomial with the number of parameters equal to the number of data points will give the exact fit. Since the two requirements of adequacy to p_t and roughness of $\widehat{p}_t$ may conflict, the choice of $\widehat{p}$ is inevitably a compromise solution.

In this paper, we discuss the penalized least squares method of smoothing a time series. The basic idea dates back at least as far as [4] and the method has been highly studied since. The paper is organized as follows. In Section 2, we present our basic method and show how a valid time series smoother can be obtained by balancing goodness of fit and lack of smoothness. Section 3 discusses the detection of outliers in hourly time series of electricity prices. The alternative method proposed by [3] is briefly outlined. The case studies in Section 3 serve to exemplify how the two methods can be used as a data preparation tool. The final section discusses our results and points out some further applications and improvements.

2 Whittacker smoother

The scope of this section is to model p_t by applying a baseline time series which best matches the original whilst being as smooth as possible. By assigning the proportion in which an increase in the goodness of fit is to be taken as counterbalancing a decrease in the smoothness, we can determine the series which best harmonizes the two requirements. The Whittacker smoother (WS) requires

$$\min_{\widehat{p}_t} Q_t\left(\widehat{p}_t\right) = L_t\left(\widehat{p}_t\right) + \lambda R_t\left(\widehat{p}_t\right) \tag{2}$$

$$L_t\left(\widehat{p}_t\right) = \sum_{t=1}^{n} \left(\widehat{p}_t - p_t\right)^2, \qquad R_t\left(\widehat{p}_t\right) = \sum_{t=m}^{n} \left[\nabla^m \widehat{p}_t\right]^2, \qquad t = 1, 2, \cdots, n \tag{3}$$

Here m is a positive integer, $\lambda > 0$ is a scalar and $\nabla = (1-B)$ is the difference operator: $\nabla \widehat{p}_t = \widehat{p}_t - \widehat{p}_{t-1}$. Note that $R_t(\widehat{p}_t)$ is null if $\widehat{p}_t$ is a polynomial of degree m.

The component L_t reflects the lack of fit, which is measured as the usual sum of squares of differences. The component R_t expresses the lack of smoothness in terms of the m-th differences of $\widehat{p}_t$. The positive constant λ controls the solution trade-off between smoothness and fidelity to p_t. If $\lambda \to 0$, then the dominant component of (2) will be the residual sum of squares and $\widehat{p}_t$ will resemble the original data points increasingly closely, no matter how irregular it may be. As the value of λ approaches ∞, the resulting baseline will approach a polynomial of degree m. Thus, an optimal relationship between the two components could and should be provided. It may be easily shown that

$$\nabla^m \widehat{p}_j = \sum_{j=i}^{n} D_{ij} \widehat{p}_j \tag{4}$$

$$D_{ij} = \begin{cases} (-1)^{m+j-i} \binom{m}{j-i} & i=1,2,\cdots,m;\, j=i,i+1,\cdots,i+m \\ 0 & \text{otherwise} \end{cases} \quad \text{for } 1 \le i \le n-m\,. \tag{5}$$

Thanks to this relationship, model (2) can be expressed in matrix notation

$$\mathbf{Q}(\widehat{\mathbf{p}}) = (\widehat{\mathbf{p}} - \mathbf{p})^t (\widehat{\mathbf{p}} - \mathbf{p}) + \lambda (\mathbf{D}\widehat{\mathbf{p}})^t (\mathbf{D}\widehat{\mathbf{p}})\,. \tag{6}$$

where $\mathbf{D}$ is a rectangular matrix with $(n-m)$ rows and n columns. Deriving $\mathbf{Q}(\widehat{\mathbf{p}})$ with respect to $\widehat{\mathbf{p}}$ and equating the partial derivatives to zero gives the linear system

$$\mathbf{A}\widehat{\mathbf{p}} = \mathbf{p} \quad \text{wih } \mathbf{A} = (\mathbf{I}_n + \lambda \mathbf{B}), \quad \mathbf{B} = \mathbf{D}^t\mathbf{D} \tag{7}$$

Here $\mathbf{I}_n$ is the $(n \times n)$ identity matrix and $\mathbf{A}$ is a symmetrical, positive definite matrix of order $(n \times n)$. It follows that the minimum occurs when $\widehat{\mathbf{p}} = \mathbf{A}^{-1}\mathbf{p}$.

Taking into account the considerable length of time series collected in the electricity market, the solution of (7) appears problematic. However, the difficulties are fewer than at first appear. Several authors, in fact, have devised very efficient computing software by exploiting the characteristics of the matrices involved. See [2]. It should be pointed out that any single value of λ could capture the smooth components of time series such as those encountered in the electricity market. A virtually flat baseline using a large value of λ will generate more false rejections of valid prices and a peak-rich one using a small value of λ will lead to false acceptances. According to [1], an objective value of λ can be obtained by minimizing the generalized cross-validation (GCV) score

$$\min_{\lambda} gcv(\lambda) = \frac{n^{-1} \sum_{t=1}^{n} [p_t - \widehat{p}_t(\lambda)]^2}{[1 - n^{-1} Tr(\mathbf{A}^{-1})]^2} \tag{8}$$

where Tr denotes the matrix trace. Criterion (8) is a measure of the leave-one-out mean squared prediction error. The effect of applying gcv to hourly time series is shown in Fig. 1 with different degrees of the polynomial in the smoothness component.

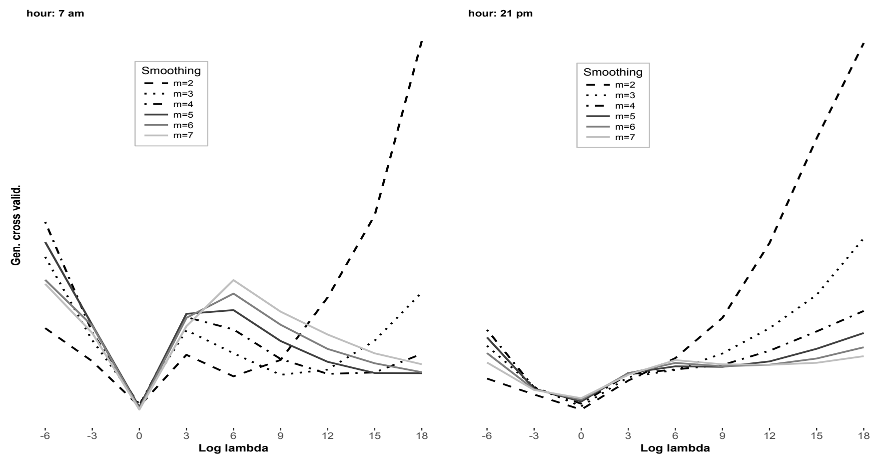

Fig. 1 Zonal electricity price in Northern Italy. Search for the best λ.

In practice, values of *gcv* are computed for many trial values of λ. Since no analytical solution exists, we need to resort to a numerical method to solve (8). We perform the minimization of (8) by performing a simple grid search for λ in a set of values for $log(\lambda) : 0, 10^j, j = 0, 1, \cdots, 9$. The $\hat{\lambda}$ value corresponding to the minimum of the minima will indicate the apparently most suitable Whittaker smoother for the time series under investigation. In so doing, the procedure for finding a baseline is fully automated. However, the simple fact that a procedure can be executed using a computer does not, of course, make it objective. It is not by coincidence that we said apparently because there is no guarantee that the λ corresponding to the global minimum will produce a valid smoother. Unfortunately, some heuristic calculations show that the minimizer of (8) will give a $\hat{\lambda}$ that leads to substantial under-smoothing and the resulting long-term pattern tends to have too many oscillations that often show up in the wrong places. These cast many doubts on the validity of (8).

As a final consideration, we note that the order of difference m is, to a limited extent, another parameter, other than λ, influencing the result of filtering a time series. However, in practice, the choice is confined to $m = 2$ or $m = 3$ because the curves for $m > 3$ show a similar pattern.

3 Detection of outliers

As stated in the introduction, we detect outliers on the basis of the difference between original and interpolated prices $\widehat{a}_t = p_t - \widehat{p}_t$ $t = 1, 2, \cdots, n$. In this regard, it is necessary to define a lower and upper bound for the residuals $\widehat{a}_t$

$$\widehat{a}_t < Q_\beta - K\left(Q_{1-\beta} - Q_\beta\right) \quad \text{or} \quad \widehat{a}_t > Q_{1-\beta} + K\left(Q_{1-\beta} - Q_\beta\right) \quad t = 1, 2, \cdots, n \quad (9)$$

where Q_β is the β-quantile and K is a positive multiplicative factor. If a residual surpasses the fences, then the corresponding price is considered an outlier. Obviously, different definitions of K and β may lead to quite different results and identification of price spikes. As an illustration of the potential merit/demerit balance of the proposed method, we compare the results of WS with those obtained by using the *tsoutliers* routine of *forecast* package in *R* software. Put briefly, *tsoutliers* is a recursive filter that decomposes the time series into seasonal, trend and irregular components, which are used to construct a reference time series based on locally weighted polynomial regression (loess). The *tsoutliers* function has the default values $\beta = 0.1$, $K = 2$. We have applied these bounds to the time series of hourly zonal electricity prices in northern Italy for the years 2013-2017. The method indicates 47 suspect spikes. The grid search proposes $\hat{\lambda} = 282'613$ and $m = 3$. To obtain the same number of outliers with the WS, we used $\beta = 0.25$ and $K = 2.714$. However, there are only 26 values in common with $tsoutliers$.

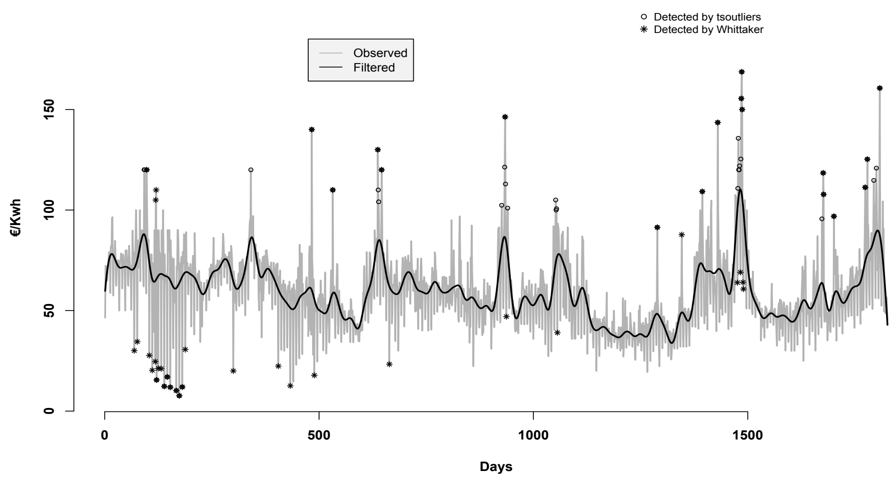

Fig. 2 Outliers detection in hourly prices, h: 10am, Zone: North, period: 2013-2017.

Once an extreme outlier has been detected and a complete time-series is required, the aberrant value may be replaced by a less unusual value. This question, however, is not discussed in the present paper.

4 Results and conclusions

In Table 1, we have collected the results obtained from the examination of $144 = 6*24$ time series (one of these has been used in Fig. 2). The data employed are hourly spot prices from the Italian day-ahead zonal market. The time series runs from 1am on Tuesday, 1st January 2013 to 24pm on Sunday, 31st December 2017, yielding data for each hour of the day and for each one of the six zones of the Italian market 1826 days. Table 1 shows the frequency of outliers and the three quartiles of the relative magnitude of outliers: $(p_t - \hat{p}_t)/p_t, p_t > 0$. The percentage of peaks and valleys that were marked as outliers by both methods is about 24‰ (in practice, 18 hourly outliers each month), which is less than a half of that detected by at least one of the filters. The findings in the table reveal that WS tends to be more aggressive than $tsoutliers$ because it not only locates a greater number of outliers than $tsoutliers$, but also because their relative size is smaller.

Table 1 Comparison of the Whittaker filter with the Hyndman filter.

	N	Q_1	Q_2	Q_3		N	Q_1	Q_2	Q_3
$H:$	46‰	0.110	0.318	0.496	$W:$	35‰	0.146	0.240	0.389
$H \cup W:$	56‰	0.112	0.301	0.521	$H \cap W:$	24‰	0.175	0.404	1.148

The two methods agree fairly well with each other in the case of extreme spikes, asis evident from the quartiles of $H \cap W$. Since we have little information about the real position and magnitude of the possible outliers, it is unclear which filter is more appropriate for our type of data.

In conclusion we can say that, although it is not possible to identify a single best method for outlier detection in day-ahead Italian zonal hourly electricity prices, the two procedures that we have applied provide sufficiently distinct results to allow them to be used in combination. Therefore, it seems logical to suggest that aberrant prices should be searched for among those that lie outside the fences of both methods.

References

1. Craven, P., Wahba, G.: Smoothing noisy data with spline functions estimating the correct degree of smoothing by the method of generalized cross-validation. Numer. Math. **31**, 377–403 (1979)
2. Garcia, D.: Robust smoothing of gridded data in one and higher dimensions with missing values. Comput. Stat. Data Anal. **54**, 1167–1178 (2010)
3. Hyndman, R.J. and Khandakar, Y . Automatic time series forecasting: the forecast package for R. J. Stat. Softw. **26**,1–22 (2008)
4. Whittaker, E.T.: On a new method of graduation. Proc. R. Soc. Edinburgh. **44**, 77–83 (1923)