

OUTLIER DETECTION AND ESTIMATION IN NONLINEAR TIME SERIES

BY FRANCESCO BATTAGLIA AND LIA ORFEI

Università La Sapienza

First Version received November 2002

Abstract. The problem of identifying the time location and estimating the amplitude of outliers in nonlinear time series is addressed. A model-based method is proposed for detecting the presence of additive or innovational outliers when the series is generated by a general nonlinear model. We use this method for identifying and estimating outliers in bilinear, self-exciting threshold autoregressive and exponential autoregressive models. A simulation study is performed to test the proposed procedures and comparing them with the methods based on linear models and linear interpolators. Finally, our results are applied for detecting outliers in the Canadian lynx trappings and in the sunspot numbers data.

Keywords. Bilinear models; exponential autoregressive models; outliers; self-exciting threshold autoregressive models; state-dependent models; sunspot numbers.

1. INTRODUCTION

Time series are often perturbed by occasional unpredictable events that generate aberrant observations. They may be due to gross errors arising in the measurement, collection and processing of the data, or to some unusual event influencing the analysed phenomenon, such as wars, strikes, an economic crisis or a temporary change in experimental conditions.

Outliers may have a significant impact on the results of standard methodology for time series analysis, therefore it is important to detect them, estimate their effects and undertake the appropriate corrective actions. For example, the impact of outliers on parameter estimation has been studied by Peña (1990), on autoregressive moving-average (ARMA) identification by Deutsch *et al.* (1990), and the effects on forecasts are addressed by Ledolter (1989) and Chen and Liu (1993a).

If the time location of an outlier is known, the intervention analysis (Box and Tiao, 1975) or missing value methods (e.g. Ljung, 1989; Beveridge, 1992) may be useful. The case of the unknown location is much more controversial and has attracted a considerable interest in the literature. Two frameworks may be distinguished. First, a model-based approach, where the underlying

structure of the time series is supposedly known. Starting from the paper of Fox (1972) concerning autoregressive models, this approach has been extended to ARMA models by several authors (Tsay, 1986, 1988; Chang *et al.*, 1988; Chen and Liu, 1993b). A Bayesian model-based approach was proposed by Abraham and Box (1979) for autoregressive models, by Smith and West (1983) for sequential decisions with dynamic linear systems, and more recently by McCulloch and Tsay (1994) and Barnett *et al.* (1996) both employing Markov chain Monte Carlo methods. The second framework is non-parametric and is based on the relationship between outliers and linear interpolators (Ljung, 1989, 1993; Peña, 1990; Peña and Maravall, 1991). However, the need for employing finite interpolators in practical applications makes the results of this approach substantially equivalent to those obtained for additive outliers when a sufficiently wide autoregressive model is fitted.

If more than one outlier is present in the time series, masking effects may seriously affect identification. A widely used linear model-based approach has been proposed by Chen and Liu (1993b) which consists of an iterative procedure that identifies outliers sequentially by searching for the most relevant anomaly, estimating its effect and removing it from the data, estimating again the model parameters on the corrected series, and iterating the process until no significant perturbation is found.

Only in recent years, the attention of researchers has shifted to the problem of our interest – the detection of outliers in nonlinear time series. The most relevant contributions to the analysis of nonlinear time series affected by one or more outliers deal with robust recursive estimates. In particular, for bilinear models, Gabr (1998) investigated a modification of some robustified versions of methods used in linear time series models. For threshold models, Chan and Cheung (1994) modified the class of generalized M-estimates usually applied to linear models. Finally, outlier detection in autoregressive conditionally heteroscedastic (ARCH) and generalized ARCH (GARCH) models was studied by van Dijk *et al.* (1999), with particular attention to robust testing, while Chen (1997) adopted a Bayesian approach via the Gibbs sampler to detect additive outliers in bilinear models.

We believe that the choice of a suitable time series model is important when searching for outliers, because, on the one hand, a large residual variance caused by overall lack of fit would result in under-identification of outliers, while on the other, a model unable to explain the local behaviour of the series would yield single large residuals, resulting in over-identification.

The outline of the paper is as follows: Section 2 deals with the model-based formulation mentioned above, and gives a general result obtained using observed residuals, with the assumption that the observed series follows the same model as the unperturbed series; Section 3 contains results for three widely used models: bilinear, self-exciting threshold autoregressive and exponential autoregressive. Section 4 reports results of some simulations and application to two real series (Canadian lynx trappings and sunspot numbers), and Section 5 draws some conclusions.

2. OUTLIER DETECTION IN GENERAL NONLINEAR MODELS

Many authors, such as Chen and Liu (1993b) and Chan (1995), distinguish essentially four characterizations of outliers found in time series data: the additive outlier (AO), the innovational outlier (IO), the level shift (LS) and the temporary change (TC). An AO affects only the level of a particular observation while an IO affects all observations beyond a certain time point through the memory of the underlying process. A level shift is an event the effect of which becomes permanent on time series, and a temporary change is an event having an initial impact the effect of which decreases exponentially according to a fixed dampening parameter. We shall focus on the first two types (additive and innovational outliers).

We suppose that the observed phenomenon may be described by a stationary zero-mean process $\{x_t\}$ following the model:

$$x_t = f\left(x^{(t-1)}; \varepsilon^{(t-1)}\right) + \varepsilon_t$$

where f is a nonlinear function also containing unknown parameters, $x^{(t-1)} = (x_{t-1}, x_{t-2}, \dots, x_{t-p})'$, $\varepsilon^{(t-1)} = (\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-s})'$; $\{\varepsilon_t\}$ is a zero-mean Gaussian white-noise series with $E(\varepsilon_t^2) = \sigma^2$.

The observed data (y_1, \dots, y_n) are realizations of a perturbed process with an outlier at time q , ($1 < q < n$), defined as follows:

$$y_t = x_t + \omega_q \delta_{t,q}, \quad \delta_{t,q} = \begin{cases} 1 & \text{if } t = q \\ 0 & \text{if } t \neq q \end{cases}$$

if the outlier is additive, and

$$y_t = f\left(y^{(t-1)}; \eta^{(t-1)}\right) + \eta_t, \quad \eta_t = \varepsilon_t + \omega_q \delta_{t,q} \tag{1}$$

where $\eta^{(t-1)} = (\eta_{t-1}, \eta_{t-2}, \dots, \eta_{t-s})'$ and $y^{(t-1)} = (y_{t-1}, y_{t-2}, \dots, y_{t-p})'$, if the outlier is of an innovational type.

The quantity ω_q is assumed constant and unknown, and we shall refer to it as the amplitude of the outlier.

We believe that the iterative framework of Chen and Liu (1993b) may be quite effective in the nonlinear case also, where masking effects can be even more serious. Thus we shall adopt a similar strategy, based on the following steps:

- Step 1. Derive initial estimates of the model parameters.
- Step 2. Given the parameter values, for any q and for each type of outliers, assume that an outlier has occurred at time q , and estimate its amplitude. If the largest absolute estimated amplitude is significant (i.e. larger than an *a priori* fixed sensitivity level, usually 3.5 or 4 times its estimated standard error), identify an outlier of that type at that time; otherwise stop.
- Step 3. Remove the effect of the identified outlier by subtracting its estimated amplitude from y_q (and also correcting all subsequent observations according to the estimated model in case of innovational outlier).

Step 4. Estimate again the model parameters on the corrected series, and iterate step 2.

While step 3 requires at most a limited computational effort, the core of the procedure lies in step 2, and we shall concentrate on it.

In order to estimate ω_q , we adopt a conditional maximum likelihood approach, and assume that, given the model and the parameter values, the likelihood function of the data is proportional to the likelihood of the residuals. For any given q , we shall assume that the function f and its parameters are known, that an outlier (of either type) has occurred at time q , and derive an estimate of its amplitude $\hat{\omega}_q$ by maximizing the conditional likelihood with respect to ω_q . In order to compute it, we need initial values. We shall suppose that no outlier occurs for $t = 1, 2, \dots, r = \max(p, s)$, and estimate ω_q only for $q > r$, therefore we need x_r, \dots, x_{r-p} , which are actually observed, and $\varepsilon_r, \dots, \varepsilon_{r-s}$ that may be obtained from $\varepsilon_t = x_t - f(x^{(t-1)}; \varepsilon^{(t-1)})$ setting $\varepsilon_k = 0, x_k = 0$ for $k \leq 0$. Conditional on these initial values, we may write the likelihood as

$$l(y|\omega) = l(x) \propto l(\varepsilon) = (2\pi\sigma^2)^{-(n-r)/2} \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=r+1}^n (\varepsilon_t)^2\right\} \quad (2)$$

Therefore, we shall estimate ω_q by minimizing the sum of squares $\Sigma \varepsilon_t^2$. While (2) holds exactly for linear models, it may only be considered as an approximation for nonlinear models (Priestley, 1988).

Given the data, the model and its parameters, and conditional on the initial values, we can compute the observed residuals

$$\eta_t = y_t - f\left(y^{(t-1)}; \eta^{(t-1)}\right), \quad t = r + 1, \dots, n$$

which, when no outlier is present, coincide with ε_t . If an innovational outlier with amplitude ω_q occurs at time q , then from (1)

$$\eta_t = \varepsilon_t, \quad t < q; \quad \eta_q = \varepsilon_q + \omega_q; \quad \eta_{q+j} = \varepsilon_{q+j}, \quad j = 1, \dots, n - q$$

thus

$$\sum_{t=r+1}^n \varepsilon_t^2 = \sum_{t=r+1}^{q-1} \eta_t^2 + (\eta_q - \omega_q)^2 + \sum_{t=q+1}^n \eta_t^2 \quad (3)$$

and minimization with respect to ω_q yields $\hat{\omega}_q = \eta_q$.

The likelihood ratio test statistic for the hypothesis H_0 : no outlier at $t = q$ against the alternative H_1 : innovational outlier at $t = q$ (following e.g. Chang *et al.*, 1988) is $\hat{\omega}_q / \hat{\sigma}_{l,q}$, where

$$\hat{\sigma}_{l,q}^2 = \frac{\eta_{r+1}^2 + \dots + \eta_{q-1}^2 + \eta_{q+1}^2 + \dots + \eta_n^2}{n - r}.$$

The statistic $\hat{\omega}_q/\hat{\sigma}_{I,q}$ is asymptotically $N(0,1)$ distributed under H_0 . Thus we adopt $\hat{\sigma}_{I,q}$ as an estimate of the standard error of the innovational outlier at time q .

Derivation of estimates in case of an additive outlier at time q is more difficult: although the observed residuals η_t are equal to ε_t only for $t < q$, and $\eta_q = \varepsilon_q + \omega_q$ as before, now the difference between η_{q+j} and ε_{q+j} , for $j > 0$, depends on ω_q and f , because each η_{q+j} is a nonlinear function of the previous η_t .

To overcome such difficulties, we consider a locally linear approximation of f obtained by a first-order Taylor expansion about $(y^{(t-1)}, \eta^{(t-1)})$:

$$f(x^{(t-1)}; \varepsilon^{(t-1)}) \cong f(y^{(t-1)}; \eta^{(t-1)}) + \sum_{j=1}^p (x_{t-j} - y_{t-j})\lambda_j(t) + \sum_{j=1}^s (\varepsilon_{t-j} - \eta_{t-j})\mu_j(t) \tag{4}$$

where

$$\lambda_j(t) = \frac{\partial}{\partial(y_{t-j})}f(y^{(t-1)}; \eta^{(t-1)}), \quad j = 1, \dots, p; \quad \lambda_j(t) = 0, \quad j > p \tag{5}$$

$$\mu_j(t) = \frac{\partial}{\partial(\eta_{t-j})}f(y^{(t-1)}; \eta^{(t-1)}), \quad j = 1, \dots, s; \quad \mu_j(t) = 0, \quad j > s \tag{6}$$

From (4) we obtain:

$$\eta_{q+j} - \varepsilon_{q+j} \cong - \left\{ \omega_q \lambda_j(q+j) + \sum_{k=1}^j (\eta_{q+j-k} - \varepsilon_{q+j-k}) \mu_k(q+j) \right\}, \quad j = 1, 2, \dots \tag{7}$$

Now, recalling that $\eta_q - \varepsilon_q = \omega_q$, and defining the recursion $c_0 = 1$,

$$c_j = - \left[\lambda_j(q+j) + \sum_{k=1}^j c_{j-k} \mu_k(q+j) \right], \quad j = 1, 2, \dots$$

we can write

$$\eta_{q+j} - \varepsilon_{q+j} \cong c_j \omega_q, \quad j = 0, 1, 2, \dots \tag{8}$$

Thus the sum of squares (3) can be rewritten:

$$\begin{aligned} \sum_{t=r+1}^n (\varepsilon_t)^2 &\cong \sum_{t=r+1}^{q-1} (\eta_t)^2 + (\eta_q - \omega_q)^2 + \sum_{t=q+1}^n [\eta_t - (\eta_t - \varepsilon_t)]^2 \\ &= \sum_{t=r+1}^{q-1} (\eta_t)^2 + \sum_{j=0}^{n-q} (\eta_{q+j} - c_j \omega_q)^2 \end{aligned}$$

and is minimized by

$$\hat{\omega}_q = \frac{\sum_{j=0}^{n-q} c_j \eta_{q+j}}{\sum_{j=0}^{n-q} c_j^2} \tag{9}$$

The estimate (9) is a weighted average of the observed residuals from $t = q$ onwards, as it happens for linear models. Here, however, the weights may be complicated functions of the parameters, the observations and the true residuals.

The likelihood ratio test statistic for the hypothesis H_0 : no outlier at $t = q$ against the alternative H_1 : additive outlier at $t = q$, turns out to be

$$\hat{\omega}_q = \frac{\sum_{j=0}^{n-q} c_j \eta_{q+j}}{\sigma_q^* \sqrt{\sum_{j=0}^{n-q} c_j^2}}$$

where

$$\sigma_q^{*2} = \frac{1}{n-r} \left\{ \sum_{t=r+1}^n \eta_t^2 - \hat{\omega}_q^2 \sum_{j=0}^{n-q} c_j^2 \right\}$$

The statistic $\hat{\omega}_q$ is asymptotically $N(0,1)$ under H_0 . Thus we assume σ_q^* as an estimate of the standard error of the additive outlier at time q .

In Section 3 we take into account specific kinds of models (bilinear, threshold and exponential autoregressive) by specifying the functional form of f . This enables us to express differences $\eta_{q+j} - \varepsilon_{q+j}$ in a more precise form and to derive improved model-based estimates of the additive outlier amplitudes.

3. OUTLIERS IN BILINEAR, THRESHOLD AND EXPONENTIAL AUTOREGRESSIVE MODELS

An expansion similar to (4) leads to a general class of nonlinear models called state-dependent models (SDM) introduced by Priestley, that may be formulated as follows (Priestley, 1988, p. 93):

$$x_t + \sum_{j=1}^p \phi_j(z^{(t-1)})x_{t-j} = \varepsilon_t + \sum_{j=1}^s \psi_j(z^{(t-1)})\varepsilon_{t-j} \tag{10}$$

where $z^{(t)}$ is the *state-vector*, i.e. $z^{(t)} = (x_{t-p+1}, \dots, x_t, \varepsilon_{t-s+1}, \dots, \varepsilon_{t-1}, \varepsilon_t)$, and $\phi_j(\cdot)$, $\psi_j(\cdot)$ correspond to the derivatives (5) and (6) respectively.

This formulation, generally known as an SDM of order (p, s) , can be interpreted as a locally linear ARMA model in which the evolution of the process at time $(t - 1)$ is governed by a set of AR coefficients $\{\phi_j(z^{(t-1)})\}$ and a set of MA coefficient $\{\psi_j(z^{(t-1)})\}$, all of which depend on the ‘state’ of the process at time point $(t - 1)$. It is also assumed that (10) satisfies the standard conditions usually imposed to identify an ARMA model.

All the models we consider (bilinear, threshold and exponential autoregressive) belong to the class of SDM models. Choosing particular analytical forms for the coefficients, it can be easily seen that SDM (10) includes the linear ARMA model, the bilinear, the self-exciting threshold autoregressive and the exponential autoregressive models as special cases.

For example, setting $\{\phi_j(z^{(t-1)})\}$ and $\{\psi_j(z^{(t-1)})\}$ as constants (i.e. independent of $z^{(t-1)}$), (10) reduces to the usual ARMA(p, s) model.

To obtain the general discrete time bilinear model, take $\{\phi_j(z^{(t-1)})\}$ as constants, α_j ,

$$\psi_j(z^{(t-1)}) = \gamma_j + \sum_{i=1}^m \beta_{ij}x_{t-i};$$

thus (10) becomes

$$x_t = \sum_{j=1}^p \alpha_j x_{t-j} + \sum_{i=1}^m \sum_{j=1}^l \beta_{ij} x_{t-i} \varepsilon_{t-j} + \sum_{j=1}^s \gamma_j \varepsilon_{t-j} + \varepsilon_t \tag{11}$$

usually indicated as BL(p, s, m, l).

Setting

$$\psi_j(z^{(t-1)}) \equiv 0 \quad \forall j \quad \text{and} \quad \phi_j(z^{(t-1)}) = \phi_j^{(i)} \quad \text{if } x_{t-d} \in R^{(i)},$$

where $\{R^{(i)}, i = 1, \dots, h\}$ is a given partition of the real numbers, d is the delay parameter and x_{t-d} is the threshold variable, (10) reduces to the SETAR($h; p$) model

$$x_t = \phi_0^{(i)} + \sum_{j=1}^p \phi_j^{(i)} x_{t-j} + \varepsilon_t^{(i)}, \quad \text{if } x_{t-d} \in R^{(i)} \tag{12}$$

Finally, in (10) take

$$\psi_j(z^{(t-1)}) \equiv 0 \quad \forall j \quad \text{and} \quad \phi_j(z^{(t-1)}) = \phi_j + \pi_j e^{-\gamma x_{t-1}^2}, \quad j = 1, \dots, p;$$

then we obtain the exponential autoregressive model of order p , defined as EXPAR(p):

$$x_t = \sum_{j=1}^p \left(\phi_j + \pi_j e^{-\gamma x_{t-1}^2} \right) x_{t-j} + \varepsilon_t \tag{13}$$

We derive specific results for each of these three models in the Appendix.

4. APPLICATIONS TO REAL AND SIMULATED SERIES

In order to evaluate the performance of the proposed procedure, some series have been simulated. The noise process was generated using the SCA system (Liu and

Hudak, 1992), under the assumption of standard normal distribution; 1000 elements were simulated and the series values were computed according to the selected model, retaining only the last 500 observations to eliminate the influence of the initial values. Then, each series was perturbed with a single additive or innovational outlier at a randomly selected time point, with amplitude 3.5 or 5.0. While the latter is generally easily recoverable by graph inspection, the first case is hardly noticeable at first sight; an example is in Figure 1, where outliers are highlighted by triangles.

The generating process was repeated 100 times independently for each of the following three models:

- i. Bilinear BL(1, 0, 1, 1) model: $x_t = 0.4x_{t-1} + 0.4x_{t-1}\varepsilon_{t-1} + \varepsilon_t$.
- ii. Self-exciting threshold SETAR(2; 1) model with threshold value equal to 1:

$$x_t = \begin{cases} 0.4 - 0.6x_{t-1} + \varepsilon_t^{(1)} & \text{if } x_{t-1} \leq 1 \\ -0.2 + 0.8x_{t-1} + \varepsilon_t^{(2)} & \text{if } x_{t-1} > 1 \end{cases}$$

- iii. Exponential autoregressive model EXPAR(2) (see Haggan and Ozaki, 1981):

$$x_t = (1.95 + 0.23e^{-x_{t-1}^2})x_{t-1} + (-0.96 - 0.24e^{-x_{t-1}^2})x_{t-2} + \varepsilon_t$$

The model parameters were estimated by means of least squares, and were generally fairly correct showing a small bias. The simulated series exhibit rather strong nonlinear features, and the outlier identification methods based on linear autoregressive models, or linear interpolators, yield unsatisfactory results.

The results of the suggested procedures are reported in Table I. For each of the 100 replications, the most significant estimated amplitude was recorded, and Table I reports how many times it corresponds to the actual location and type of the outlier. Moreover, the average estimated amplitude and its standard error, computed on the correctly identified cases, appears in the table. In all cases when the location was correctly identified, the likelihood ratio test statistics were larger

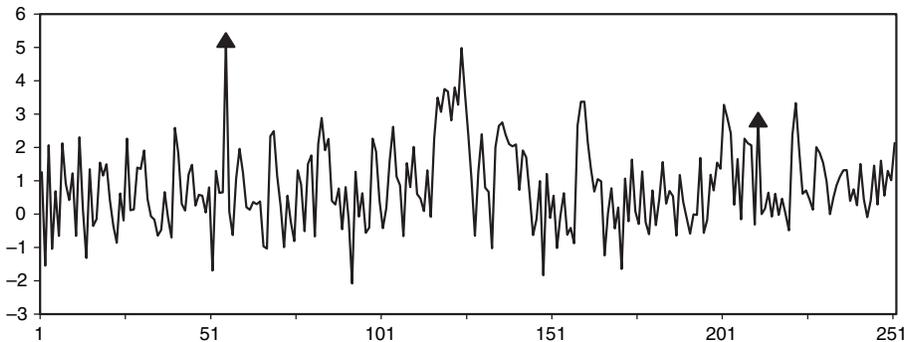


FIGURE 1. Simulated series following a SETAR(2;1) model, perturbed with additive outliers at $t = 55(\omega = 5.0)$, and at $t = 211(\omega = 3.5)$.

TABLE I
OUTLIER IDENTIFICATION ON SIMULATED SERIES

Model	Outlier type	Correct time identification	Correct type identification	Estimate average (SE)	l.r.t. > 3.5.
BILINEAR	Additive				
	$\omega = 5$	96	96	4.76 (.90)	96
	$\omega = 3.5$	88	88	3.37 (.70)	85
	Innovational				
	$\omega = 5$	91	91	4.88 (.94)	86
	$\omega = 3.5$	50	50	4.04 (.68)	39
SETAR	Additive				
	$\omega = 5$	92	85	4.98 (.94)	83
	$\omega = 3.5$	57	49	3.93 (.66)	51
	Innovational				
	$\omega = 5$	87	81	5.07 (.91)	79
	$\omega = 3.5$	44	41	4.26 (.60)	33
EXPAR	Additive				
	$\omega = 5$	99	99	5.02 (.38)	99
	$\omega = 3.5$	97	96	3.50 (.48)	94
	Innovational				
	$\omega = 5$	82	82	5.27 (.93)	75
	$\omega = 3.5$	47	47	4.28 (.79)	29

Notes: For each simulated model and contamination, the frequency of correct time and type detection on 100 replications is shown, together with the average and standard error of the estimated outlier amplitude in the correctly identified cases, and the number of times that the likelihood ratio test statistic (l.r.t.) was found larger than 3.5 (it was always larger than 3.0).

than 3.0 (high sensitivity) and often larger than 3.5 (medium sensitivity). The number of replications where the test statistic was larger than 3.5 is reported in the last column of Table I.

We found that when the perturbation is relatively large ($\omega = 5$) the procedure is very effective, both in the additive and innovational case. Moreover, the estimated amplitude shows a small bias. On the contrary, when the outliers are less relevant ($\omega = 3.5$), their detection is much more difficult (although generally correct in about one half of replications), and an innovational outlier is often less clearly identified than an additive one, especially in the SETAR case. Furthermore, the estimated amplitudes have shown a positive bias.

In order to complete our analysis, we also considered two classical examples, the Canadian lynx series and the sunspot series, and applied the proposed iterative procedure. These data sets have been widely investigated in literature, and many models have been adapted to both series; we have considered some of them and the parameter estimates proposed in literature.

For the Canadian lynx series after a logarithmic transformation, we searched for outliers using a BL(12, 0, 9, 9) as in Gabr and Subba Rao (1981); a SETAR (2; 7, 2), as in Tong and Lim (1980), and an exponential model of order 11, as in Haggan and Ozaki (1981). In all cases, the parameter values were those reported by the related papers. It is a fairly regular series and neither of these models suggests any outlier, since the likelihood test statistic values for each time are always smaller than 3 in modulus. Using a linear interpolator or a linear

TABLE II
 OUTLIER DETECTION IN THE SUNSPOT SERIES, USING MODELS SUGGESTED BY PRIESTLEY (1981)

Model	$\hat{\sigma}^2$	q	Type	$\hat{\omega}_q$	l.r.t. statistic
BL(3,0,3,4)	182	75	AO	21.14	4.11
		78	IO	51.54	3.96
SETAR(2;4,10)	171	149	AO	48.73	4.57
AR(9)	188	171	AO	36.54	4.62
		78	IO	56.07	4.51
		218	AO	29.88	4.27
		137	IO	42.0	3.58

Notes: For each fitted model, the table displays the observed residual variance, the identified outlier time (q) and type, the estimated amplitude and the observed value of the likelihood ratio test statistic.

autoregressive model, we found a slightly worse fit and likelihood ratio test statistic values between 3.0 and 3.5 at $t = 50$ (AO), $t = 97$ (IO) and $t = 16$ (AO).

For the sunspot series, we used 216 observations from the year 1700 to 1915 and considered a BL(3, 0, 3, 4), a SETAR(2; 4, 10) and an AR(9), as reported in Priestley (1981, pp. 882–7), results are shown in Table II where the iterative procedure is used and the sensitivity level is set at 3.5.

It may be seen that each method suggests the presence of outliers at different locations. The events detected using a bilinear model relate to the lower part of the cycle, while the SETAR model relates to a peak. The procedure based on linear autoregressive models seems to produce an over-identification, possibly because of the nonlinear features of this series. Thus we can draw the conclusion that some perturbation probably occurred but no uncontroversial outlier may be identified.

5. CONCLUDING REMARKS

The increasing interest towards nonlinear models for the analysis of important time series the behaviour of which is not well reproduced by linear models requires specific techniques for treating outliers in this context. Outliers are even more critical in nonlinear series than in the linear case, because, on the one hand, their effects may be much longer persistent, and, on the other, they may have a serious impact on parameter estimation, which is usually more difficult when nonlinear models are entertained. Therefore specific model-based methods, like those proposed here, are relevant.

In our simulation experiment, the bias induced by outliers on parameter estimation was not found to be very large, but we analysed a series of 500 observations: smaller series are certainly more seriously affected. However, the iterative detection procedure or robust estimation methods may help in reducing the bias.

When outliers are very large, the detection methods based on linear ARMA models or linear interpolators are relatively successful, but they are less useful

when the anomalies are not so evident. It would be interesting, however, although difficult, to derive optimal nonlinear interpolators based on different nonlinear models and study their effectiveness in outlier identification.

Finally, we can observe that various different models have the purpose of reproducing and explaining similar nonlinear features; therefore, often several different nonlinear models are fitted to the same data: we think that this may also be convenient when searching for outliers. In fact, performing an outlier detection based on linear and on some nonlinear models provides complete information on perturbed time periods, and helps to distinguish between genuinely unpredictable observations and misfitting because of the model global or local inadequacy.

APPENDIX

We derive here the explicit form of the outlier estimates for bilinear, threshold and exponential autoregressive models.

BILINEAR MODELS

Consider a series $\{y_t\}$, following model (11), which contains an AO at a generic time point q , i.e. $y_q = x_q + \omega_q$; for notational simplicity we put $\alpha_j = 0$ for $j > p$, $\beta_{ij} = 0$ for $i > m$ or $j > l$, and $\gamma_j = 0$, $j > s$. From (5) and (6), we can write:

$$\lambda_r(q+k) = \alpha_r + \sum_{j=1}^l \beta_{rj} \eta_{q+k-j}; \quad \mu_r(q+k) = \sum_{i=1}^m \beta_{ir} y_{q+k-i} + \gamma_r$$

so that the generic coefficient c_k is

$$c_k = -\alpha_k - \sum_{j=1}^l \beta_{kj} \eta_{q+k-j} - \sum_{r=1}^k c_{k-r} \left(\sum_{i=1}^m \beta_{ir} y_{q+k-i} + \gamma_r \right) \tag{A.1}$$

where, as before, $c_0 = 1$. Note that $\lambda_r(q+k) = 0$ if $r > \max(m, p)$ and $\mu_r(q+k) = 0$ if $r > \max(l, s)$. The approximate difference between the residuals (7) is:

$$\eta_{q+k} - \varepsilon_{q+k} \cong c_k \omega_q = -\omega_q \alpha_k - \omega_q \sum_{j=1}^l \beta_{kj} \eta_{q+k-j} - \sum_{r=1}^k \sum_{i=1}^m (\beta_{ir} y_{q+k-i} + \gamma_r) (\eta_{q+k-r} - \varepsilon_{q+k-r})$$

However, we may evaluate the *exact* difference as follows:

$$\begin{aligned} \eta_{q+k} - \varepsilon_{q+k} &= y_{q+k} - f\left(y^{(q+k-1)}, \eta^{(q+k-1)}\right) - x_{q+k} + f\left(x^{(q+k-1)}, \varepsilon^{(q+k-1)}\right) \\ &= -\omega_q \alpha_k - \omega_q \sum_{j=1}^l \beta_{kj} \varepsilon_{q+k-j} - \sum_{r=1}^k \sum_{i=1}^m (\beta_{ir} y_{q+k-i} + \gamma_r) (\eta_{q+k-r} - \varepsilon_{q+k-r}) \end{aligned} \tag{A.2}$$

The error consists in substituting the unperturbed residuals ε_t with the observed residuals η_t in the second term of (A.2) for $t = q+k-l, q+k-l+1, \dots, q+k-1$: if an

additive outlier occurs at time q , generally $|\eta_{q+j}| > |\varepsilon_{q+j}|$ for $j \geq 0$ and a better approximation may be achieved by deriving an estimate of the ε_{q+j} based on an initial estimate of ω_q .

Thus we compute initial coefficients c_k^* as follows, according to (A.1):

$$c_k^* = - \left[\alpha_k + \sum_{j=k+1}^l \beta_{kj} \eta_{q+k-j} + \sum_{r=1}^k c_{k-r}^* \left(\sum_{i=1}^m \beta_{ir} y_{q+k-i} + \gamma_r \right) \right]$$

and obtain a first estimate of the outlier amplitude, ω_q^* , according to the linear combination (9).

Now, through ω_q^* , we can compute approximate values ε_j^* of the unknown ε s using (11) applied to $x_t^* = x_t - \omega_q^* \delta_{t,q}$, and finally new values of c_k given by

$$\hat{c}_k = - \left[\alpha_k + \sum_{j=1}^l \beta_{kj} \varepsilon_{q+k-j}^* + \sum_{r=1}^k \hat{c}_{k-r} \left(\sum_{i=1}^m \beta_{ir} y_{q+k-i} + \gamma_r \right) \right]$$

so that the quantities $\hat{c}_k \cdot \omega$ are closer to the exact differences $\eta_{q+k} - \varepsilon_{q+k}$ given in (A.2), and a final estimate of the outlier amplitude is:

$$\hat{\omega}_q = \frac{\sum_{j=0}^{n-q} \hat{c}_j \eta_{q+j}}{\sum_{j=0}^{n-q} \hat{c}_j^2}$$

SETAR MODELS

For clarity of exposition, in (12) we shall write $\phi_j^{(i)} = \phi_j^{(x_{t-d})}$ if $x_{t-d} \in R^{(i)}$, since the regime at time t depends upon the value of the threshold variable at time $t - d$. The analogous model for the perturbed residuals is:

$$\eta_t = y_t - \phi_0^{(y_{t-d})} - \sum_{j=1}^p \phi_j^{(y_{t-d})} y_{t-j}$$

where, as before, $y_q = x_q + \omega_q$. Note that an outlier occurring at time point q may influence the correct regime identification at time $q + d$. Following the same scheme used for bilinear models, (5) and (6) become:

$$\begin{aligned} \lambda_j(q+k) &= \phi_j^{(y_{q+k-d})}, \quad j = 1, \dots, p, \quad k > 0; \\ \lambda_j(q+k) &= 0, \quad j > p, \quad k > 0; \\ \mu_j(q+k) &= 0, \quad \forall j, \quad k > 0 \end{aligned}$$

and the coefficients c_j in (8) are $c_j = -\phi_j^{(y_{q+j-d})}$. The difference between residuals is now exactly:

$$\eta_{q+j} - \varepsilon_{q+j} = \phi_0^{(x_{q+j-d})} - \phi_0^{(y_{q+j-d})} + \sum_{k=1}^p \phi_k^{(x_{q+j-d})} x_{q+j-k} - \sum_{k=1}^p \phi_k^{(y_{q+j-d})} y_{q+j-k}, \quad j = 1, \dots, p \tag{A.3}$$

If $j \neq d$, x_{q+j} and y_{q+j} belong to the same regime, i.e. the coefficients of the two series coincide: therefore the only term different from zero in (A.3) is that for $k = j$: $\eta_{q+j} - \varepsilon_{q+j} = -\phi_j^{(y_{q+j-d})} \omega_q$. On the contrary, if $j = d$, the regime may change, because the outlier affects the threshold variable; (A.3) becomes:

$$\eta_{q+d} - \varepsilon_{q+d} = \phi_0^{(x_q)} - \phi_0^{(y_q)} + \sum_{k=1}^p \phi_k^{(x_q)} x_{q+d-k} - \sum_{k=1}^p \phi_k^{(y_q)} y_{q+d-k} \tag{A.4}$$

The second difference on the right-hand side of (A.4) contains the perturbed term if $d = k$. This case is possible only if $d \leq p$; thus, assuming that $d \leq p$,

$$\eta_{q+d} - \varepsilon_{q+d} = \phi_0^{(x_q)} - \phi_0^{(y_q)} + \sum_{k \neq d} y_{q+d-k} (\phi_k^{(x_q)} - \phi_k^{(y_q)}) + y_q (\phi_d^{(x_q)} - \phi_d^{(y_q)}) - \omega_q \cdot \phi_d^{(x_q)}$$

Generally, the coefficients $\phi_j^{(x_q)}$ are not known, because they are related to the series $\{x_t\}$; one possible way to proceed is to evaluate a preliminary estimate ω_q^* considering only the differences that are not affected by the presence of unknown terms, i.e. differences for $j \neq d$ in (9), or alternatively putting $c_d = -\phi_d^{(y_q)}$; we then estimate x_q by analogy through $x_q^* = y_q - \omega_q^*$ and compute a modified residual η_{q+d}^* as follows:

$$\eta_{q+d}^* = \eta_{q+d} - (\phi_0^{(x_q^*)} - \phi_0^{(y_q)}) - \sum_{k \neq d} y_{q+d-k} (\phi_k^{(x_q^*)} - \phi_k^{(y_q)}) - y_q (\phi_d^{(x_q^*)} - \phi_d^{(y_q)})$$

thus to a closer approximation $\eta_{q+d}^* - \varepsilon_{q+d} = c_d^* \omega_q$, where $c_d^* = -\phi_d^{(x_q^*)}$.

Letting $\eta_{q+j}^* = \eta_{q+j}$ and $c_j^* = c_j$ for $j \neq d$, the final estimate is obtained as

$$\hat{\omega}_q = \frac{\eta_q^* + \sum_{j=1}^p c_j^* \eta_{q+j}^*}{1 + \sum_{j=1}^p c_j^{*2}}$$

We note finally that in the somewhat unusual case that $d > p$ the possible regime change at $t = q$ does not affect the coefficients c_j , $j = 1, \dots, p$; therefore, a two-step estimation is not required.

EXPONENTIAL AUTOREGRESSIVE MODELS

The function in (13) depends only upon observations, so that the first derivatives with respect to the residuals are zero for every t ; (5) becomes, in this case:

$$\begin{aligned} \lambda_1(q+k) &= \phi_1 + \pi_1 e^{-\gamma y_{q+k-1}^2} - 2\gamma \pi_1 \sum_{j=1}^p y_{q+j-1} e^{-\gamma y_{q+k-1}^2}; \\ \lambda_j(q+k) &= \phi_j + \pi_j e^{-\gamma y_{q+k-1}^2}, \quad j = 2, \dots, p \end{aligned}$$

so the coefficients c_j coincide with $-\lambda_j(q + j)$. The difference between residuals is:

$$\begin{aligned} \eta_{q+1} - \varepsilon_{q+1} &= \sum_{j=1}^p \left\{ (\phi_j + \pi_j e^{-\gamma x_q^2}) x_{q+1-j} - (\phi_j + \pi_j e^{-\gamma y_q^2}) y_{q+1-j} \right\} \\ &= -(\phi_1 + \pi_1 e^{-\gamma y_q^2}) \omega_q + (e^{-\gamma x_q^2} - e^{-\gamma y_q^2}) \sum_{j=1}^p \pi_j x_{q+1-j} \end{aligned}$$

$$\eta_{q+k} - \varepsilon_{q+k} = -\left(\phi_k + \pi_k e^{-\gamma y_{q+k-1}^2}\right)\omega_q, \quad k = 2, \dots, p$$

The first expression contains a second term, that we shall denote by ζ , which depends upon the amount x_q , that is generally unknown. We can proceed as before, deriving a first estimate ω_q^* using only the differences not affected by unknown terms, i.e. the differences for $k = 2, \dots, p$, in (9) (or $\omega_q^* = \eta_q$, if $p = 1$) and x_q can now be estimated through $x_q^* = y_q - \omega_q^*$, and its value used to estimate the amount ζ :

$$\zeta^* = \left(e^{-\gamma x_q^{*2}} - e^{-\gamma y_q^2}\right) \left(\pi_1 x_q^* + \sum_{j=2}^p \pi_j y_{q+1-j}\right)$$

so that we can obtain the final estimate of the anomaly by

$$\hat{\omega}_q = \frac{\eta_q^* + \sum_{j=1}^p c_j \eta_{q+j}^*}{1 + \sum_{j=1}^p c_j^2}$$

where $\eta_{q+1}^* = \eta_{q+1} - \zeta^*$ and $\eta_{q+j}^* = \eta_{q+j}$, $j = 2, \dots, p$.

ACKNOWLEDGEMENTS

This work was supported by Ministero dell'Università e della Ricerca Scientifica and by Consiglio Nazionale delle Ricerche. We wish to thank an anonymous referee for valuable comments and suggestions.

NOTES

Corresponding author: Francesco Battaglia, Dipartimento di Statistica, Probabilità e Statistiche Applicate, Università La Sapienza, Piazzale Aldo Moro 5, 00100 Roma, Italy. Fax: +39-06-4959241; E-mail: francesco.battaglia@uniroma1.it

REFERENCES

- ABRAHAM, B. and BOX, G. E. P. (1979) Bayesian analysis of some outlier problems in time series. *Biometrika* 66, 229–36.
- BARNETT, G., KOHN, R. and SHEATHER S. (1996) Bayesian estimation of an autoregressive model using Markov chain Monte Carlo. *Journal of Econometrics* 74, 237–54.
- BEVERIDGE S. (1992) Least square estimation of missing values in time series. *Communications in Statistics – Theory and Methods* 21, 3479–96.
- BOX, G. E. P. and TIAO, G. C. (1975) Intervention analysis with application to environmental and economic problems. *Journal of the American Statistical Association* 70, 70–9.
- CHAN, W. S. (1995) Understanding the effect of time series outliers on sample autocorrelation. *Test* 4, 179–86.

- CHAN, W. S. and CHEUNG, S. H. (1994) On robust estimation of threshold autoregressions. *Journal of Forecasting* 13, 37–49.
- CHANG, I., TIAO, G. C. and CHEN, C. (1988) Estimation of time series parameters in the presence of outliers. *Technometrics* 30, 193–204.
- CHEN, C. (1997) Detection of additive outliers in bilinear time series. *Computational Statistics and Data Analysis* 24, 283–94.
- CHEN, C. and LIU, L. M. (1993a) Forecasting time series with outliers. *Journal of Forecasting* 12, 13–35.
- CHEN, C. and LIU, L. M. (1993b) Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association* 88, 284–97.
- DEUTSCH, S. J., RICHARDS, J. E. and SWAIN, J. J. (1990) Effects of a single outlier on ARMA identification. *Communications in Statistics – Theory and Methods* 19, 2207–27.
- VAN DIJK, D., FRANCES, P. H. and LUCAS, A. (1999) Testing for ARCH in the presence of additive outliers. *Journal of Applied Econometrics* 14, 539–62.
- FOX, A. J. (1972) Outliers in time series. *Journal of the Royal Statistical Society Series B* 34, 350–63.
- GABR, M. M. (1998) Robust estimation of bilinear time series models. *Communications in Statistics – Theory and Methods* 27, 41–53.
- GABR, M. M. and SUBBA RAO, T. (1981) The estimation and prediction of subset bilinear time series models with applications. *Journal of Time Series Analysis* 2, 155–71.
- HAGGAN, V. and OZAKI, T. (1981) Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model. *Biometrika* 68, 189–96.
- LEDOLTER, J. (1989) The effect of additive outliers on the forecasts from ARMA models. *International Journal of Forecasting* 5, 231–40.
- LIU, L. and HUDAK, G. B. (1992) *Forecasting and Time Series Analysis using the SCA Statistical System*. Oak Brook, IL: Scientific Computing Associates.
- LJUNG, G. M. (1989) A note on the estimation of missing values in time series. *Communications in Statistics – Simulation Computation* 18, 459–65.
- LJUNG, G. M. (1993) On outlier detection in time series. *Journal of the Royal Statistical Society Series B* 55, 559–67.
- MCCULLOCH, R. E. and TSAY, R. S. (1994) Bayesian analysis of autoregressive time series via the Gibbs sampler. *Journal of Time Series Analysis* 15, 235–50.
- PEÑA, D. (1990) Influential observations in time series. *Journal of Business and Economic Statistics* 8, 235–41.
- PEÑA, D. and MARAVALL, A. (1991) Interpolations, outliers and inverse autocorrelations. *Communications in Statistics – Theory and Methods* 20, 3175–86.
- PRIESTLEY, M. B. (1981) *Spectral Analysis and Time Series*. London: Academic Press.
- PRIESTLEY, M. B. (1988) *Non-linear and Non-stationary Time Series Analysis*. London: Academic Press.
- SMITH, A. F. M. and WEST, M. (1983) Monitoring renal transplants: an application of the multiprocess Kalman filter. *Biometrics* 39, 867–78.
- TONG, H. and LIM, K. S. (1980) Threshold autoregressive, limit cycles and cyclical data. *Journal of the Royal Statistical Society Series B* 42, 245–92.
- TSAY, R. S. (1986) Time series model specification in the presence of outliers. *Journal of the American Statistical Association* 81, 131–41.
- TSAY, R. S. (1988) Outliers, level shifts and variance changes in time series. *Journal of Forecasting* 7, 1–20.

Copyright of Journal of Time Series Analysis is the property of Blackwell Publishing Limited and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.