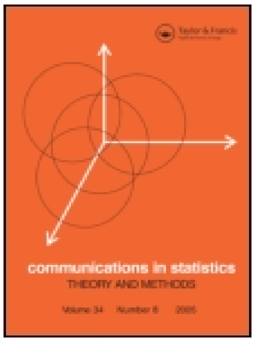

Communications in Statistics - Theory and Methods

ISSN: 0361-0926 (Print) 1532-415X (Online) Journal homepage: http://www.tandfonline.com/loi/lsta20

The Identification of Multiple Outliers in ARIMA Models

María Jesús Sánchez & Daniel Peña

To cite this article: María Jesús Sánchez & Daniel Peña (2003) The Identification of Multiple Outliers in ARIMA Models, Communications in Statistics - Theory and Methods, 32:6, 1265-1287, DOI: 10.1081/STA-120021331

To link to this article: http://dx.doi.org/10.1081/STA-120021331

MARCEL DEKKER, INC. • 270 MADISON AVENUE • NEW YORK, NY 10016

COMMUNICATIONS IN STATISTICS
Theory and Methods
Vol. 32, No. 6, pp. 1265–1287, 2003

The Identification of Multiple Outliers in ARIMA Models

María Jesús Sánchez[1,*] and Daniel Peña[2]

[1]Laboratorio de Estadística, Universidad Politécnica de Madrid, Madrid, Spain
[2]Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, Getafe, Madrid, Spain

ABSTRACT

There are three main problems in the existing procedures for detecting outliers in ARIMA models. The first one is the biased estimation of the initial parameter values that may strongly affect the power to detect outliers. The second problem is the confusion between level shifts and innovative outliers when the series has a level shift. The third problem is masking. We propose a procedure that keeps the powerful features of previous methods but improves the initial parameter estimate, avoids the confusion between innovative outliers

*Correspondence: María Jesús Sánchez, Laboratorio de Estadística, Universidad Politécnica de Madrid, José Gutierrez Abascal, 2 28006 Madrid, Spain; E-mail: mjsan@etsii.upm.es.

DOI: 10.1081/STA-120021331
0361-0926 (Print); 1532-415X (Online)

and level shifts and includes joint tests for sequences of additive outliers in order to solve the masking problem. A Monte Carlo study and one example of the performance of the proposed procedure are presented.

Key Words: Equivalent configurations; Influential observations; Mis-specifications; Robust estimation.

1. INTRODUCTION

The study of outliers in ARIMA models has been a very active field of research. Fox (1972) defines additive and innovative outliers and proposes the use of maximum likelihood ratio tests to detect them. Chang and Tiao (1983) and Chang et al. (1988) extend the results of Fox (1972) to ARIMA models and present an iterative procedure for outlier detection and parameter estimation. Tsay (1988) generalizes this procedure for detecting level shifts and temporary changes. Balke (1993) proposes a method to solve the confusion between level shift and innovative outliers. Chen and Liu (1993) present an outlier detection and parameter estimation procedure that includes several improvements over previous procedures and seems to be widely used. However, this procedure may misidentify level shifts as innovative outliers, and some outliers may not be identified due to masking effects. Outliers are not necessarily influential observations and Peña (1990, 1991) presents statistics to measure the influence of outliers on the model parameters. Bruce and Martin (1989) define two diagnostics for detecting outlier patches. The first measures the change in the estimate of the ARIMA coefficients and the second the change in the estimated variance. Other useful references for outlier detection in time series models are Bustos and Yohai (1986), McCulloch and Tsay (1993, 1994), Le et al. (1996), and Justel et al. (2001).

There are three main problems in the existing procedures for detecting outliers in ARIMA models. (a) Confusion between level shift and innovative outliers when the series has a level shift. (b) Masking, the usual procedures based on the identification of outliers one by one, may fail in the identification of patches of outliers. (c) Biased estimation of the initial parameter values: if the sample contains influential outliers, the initial parameter estimates made under the hypotheses of no outliers can be very biased and the procedure may fail.

In order to illustrate these problems, we have simulated $n = 100$ observations from the AR(1) model $(1 - \phi B)y_t = a_t$, with $\phi = 0.6$, and

$\sigma_a = 1$. In the generated sample, the parameter ϕ is estimated as $\hat{\phi} = 0.45$. To show the confusion between level shift and innovational outliers, we have introduced at $t = 50$ a level shift of size equal to 4, and at $t = 51$ an additive outlier of size 4. Figure 1 shows the plot of the contaminated time series. With the procedure proposed by Chen and Liu (1993), CL from now on, and using a critical value $C = 3$, three outliers are detected: an innovative outlier at $t = 50$ with estimated effect 3.74, a level shift at $t = 51$ with estimated effect 4.41, and an innovative outlier at $t = 98$ with estimated effect -2.78, and the autoregressive parameter is estimated at the end of the procedure as $\hat{\phi} = 0.38$. Note that the CL procedure not only misclassifies the two types of outliers, but leads to a downward-biased parameter estimate that generates a small false outlier at $t = 98$. Also, the effect on observation 51 is strongly under-evaluated (the total effect should be 8, and the model estimates 5.8), which produces a false large positive residual at this point. With the procedure proposed in this article the two outliers are correctly identified, their estimated effects are 4.26 and 3.32, respectively, and the final estimated parameter value is $\hat{\phi} = 0.42$.

To illustrate the effect of outlier patches and bad initial parameter estimates, we now introduce in the previously simulated time series two patches of outliers: two additive outliers of sizes -4 and 4 at $t = 16$ and 17, respectively, and a level shift of size 3 at $t = 40$ followed by an additive outlier of size 4 at $t = 41$. Figure 2 shows the plot of this contaminated time series. With the CL procedure and

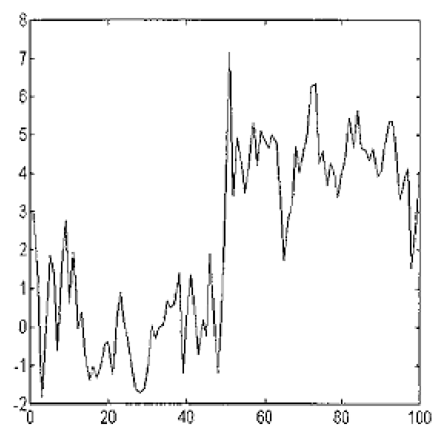

Figure 1. Plot of a simulated AR(1) series with a level shift at $t = 50$ and an additive outlier at $t = 51$.

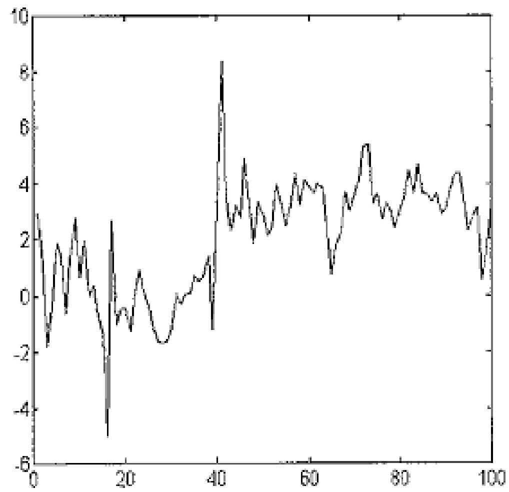

Figure 2. Plot of a simulated AR(1) series with three additive outliers at $t = 16$, 17, 41, and a level shift at $t = 40$.

using the same critical value, only an additive outlier at $t = 41$ is detected. Its estimated effect is 5.06, and the final parameter estimates are $\hat{\phi} = 0.86$ and $\widehat{\sigma}_a = 1.44$. As the level shift pulls the $\hat{\phi}$ parameter towards one, and the other outliers inflate the estimated variance, the CL procedure breaks down. Only the very large outlier at $t = 41$ is identified, and as a consequence both the AR parameter and the variance are strongly biased. With the procedure proposed in this article the four outliers are correctly identified and their estimated effects are respectively -4.31 ($t = 16$, additive outlier), 3.29 ($t = 17$, additive outlier), 3.26 ($t = 40$, level shift) and 5.04 ($t = 41$, additive outlier), with $\hat{\phi} = 0.42$ and $\widehat{\sigma}_a = 0.90$.

The rest of the article is organized as follows. Section 2 presents the model and the notation. Section 3 analyzes the confusion between innovational outliers and level shifts. Section 4 discusses the masking problem with sequences of additive outliers and presents a possible solution to this problem. Section 5 describes the proposed method which modifies the one by Chen and Liu (1993) by incorporating the results of the two previous sections as well as a robust initial parameter estimation. Section 6 studies the performance of the proposed procedure in one example. Section 7 contains some concluding remarks.

2. MODEL AND NOTATION

Let y_t be a stochastic process following an ARIMA model

$$\phi(B) \nabla^d y_t = \theta(B)a_t, \tag{1}$$

where B is the backshift operator such that $By_t = y_{t-1}$, $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$ are polynomials in B of degrees p and q, respectively, with roots outside the unit circle, $\nabla = 1 - B$ is the difference operator and a_t is a white-noise sequence of iid $N(0, \sigma_a^2)$ variables. The AR(∞) representation of the process is given by $\pi(B)y_t = a_t$, where $\pi(B) = (1 - \pi_1 B - \pi_2 B^2 - \cdots) = \theta^{-1}(B) \nabla^d \phi(B)$. In order to allow for outliers, we assume that instead of y_t we observe an outlier-contaminated series z_t given by

$$z_t = \omega_i V_i(B) I_t^{(T_i)} + y_t \tag{2}$$

where ω_i is the outlier size, $I_t^{(T_i)}$ is an impulse variable that takes the value 1 if $t = T_i$ and 0 otherwise and $V_i(B)$ defines the outlier type. When $V_i(B) = 1/\pi(B)$ we have an innovative outlier (IO), when $V_i(B) = 1$ an additive outlier (AO), and when $V_i(B) = 1/(1 - B)$ a level shift (LS). Then, calling $e_t = \pi(B)z_t$, Eq. (2) can be written as

$$e_t = \omega_i x_t + a_t \tag{3}$$

where for an IO $\omega_i = \omega_I$ and $x_t = I_t^{(T)}$, for an AO $\omega_i = \omega_A$ and $x_t = \pi(B)I_t^{(T)}$, and for a LS $\omega_i = \omega_L$ and $x_t = \pi(B)(1 - B)^{-1} I_t^{(T)}$. In order to test for a single outlier, the following hypotheses are usually considered: (a) $H_0 : \omega_I = \omega_A = \omega_L = 0$; (b) $H_I : \omega_I \neq 0$, (c) $H_A : \omega_A \neq 0$, (d) $H_L : \omega_L \neq 0$, and the likelihood ratio test statistics for testing H_0 vs. H_I, H_A, and H_L are respectively $\lambda_{i,T} = \widehat{\omega}_i / \sigma_i$ for $i = I, A$, and L, where σ_i is the standard deviation of the estimate. Under the null hypothesis of no outliers, these statistics are asymptotically distributed as $N(0, 1)$.

In practice, the model parameters are unknown. Then the parameters are initially estimated by assuming that there are no outliers and the detection is based on the statistics $\hat{\lambda}_{I,T}, \hat{\lambda}_{A,T}$, and $\hat{\lambda}_{L,T}$, in which the parameters are substituted by their estimates. These statistics are asymptotically equivalent to $\lambda_{I,T}, \lambda_{A,T}$, and $\lambda_{L,T}$, respectively. For detecting outliers at an unknown position Chang et al. (1988) and Tsay (1988) suggest calculating at each sample point the statistic

$$\eta_t = \max\{|\hat{\lambda}_{I,t}|, \quad |\hat{\lambda}_{A,t}|, \quad |\hat{\lambda}_{L,t}|\},$$

and if max $\eta_t = |\hat{\lambda}_{I,T}| \geq C$, where C is a predetermined constant, we assume an IO in $t = T$, if max $\eta_t = |\hat{\lambda}_{A,T}| \geq C$ an AO, and if max $\eta_t = |\hat{\lambda}_{L,T}| \geq C$ a LS.

3. THE CONFUSION BETWEEN INNOVATIVE OUTLIERS AND LEVEL SHIFTS

Suppose that we have a stationary AR process with a level shift (LS). As shown by Chen and Tiao (1990) the observed series would seem a random walk and the estimated AR parameter will be close to one. As a level shift (LS) and an innovational outlier (IO) are identical on a random walk, these two effects can be confused and the empirical evidence (see Balke (1993)) shows that a level shift in a stationary time series is usually identified as an innovational outlier.

In order to understand better this situation, first let us compare the statistics for testing for an IO and a LS. Calling $\hat{e}_T = \widehat{\pi}(B)z_t$ to the residuals computed from the ARIMA model where the parameters are estimated by maximum likelihood (ML) assuming no outliers, and calling $\hat{\sigma}_a$ to the ML estimate of σ_a, the statistic for testing for LS is:

$$\hat{\lambda}_{L,T} = \frac{\hat{e}_T - \sum_{i=1}^{n-T} \hat{l}_i \hat{e}_{T+i}}{\hat{\sigma}_a (1 + \sum_{i=1}^{n-T} \hat{l}_i^2)^{1/2}},$$

and using that $\hat{l}_j = -1 + \sum_{i=1}^{j} \hat{\pi}_i$, as the statistic for testing for IO is $\hat{\lambda}_{I,T} = \hat{e}_T / \hat{\sigma}_a$, both statistics are related by:

$$\hat{\lambda}_{L,T} = \frac{\hat{\lambda}_{I,T} + \dfrac{\sum_{i=1}^{n-T}(\hat{e}_{T+i}(1 - \sum_{j=1}^{i} \hat{\pi}_j))}{\hat{\sigma}_a}}{(1 + \sum_{i=1}^{n-T}(1 - \sum_{j=1}^{i} \hat{\pi}_j)^2)^{1/2}}. \tag{4}$$

Note that this equation shows that for an AR(1) model, when $\phi \to 1$ both statistics are equal, because then the two outlier models are identical. However, if $\phi < 1$ we expect $\hat{\lambda}_{L,T}$ to be smaller than $\hat{\lambda}_{I,T}$. In general, for ARMA models, the statistic for LS will be similar to the one for IO when the larger AR root approaches the non-stationary value of 1 or when $(1 - \sum_{j=1}^{i} \hat{\pi}_j)$ is small for all $i \geq 1$. However, for stationary and invertible ARMA models, when T is not close to the end of the series and n is large, the second term will go to zero and the likelihood ratio for level shifts, $\hat{\lambda}_{L,T}$, is expected to be smaller than the likelihood ratio for innovational outliers, $\hat{\lambda}_{I,T}$. This result suggests that the critical values under the null hypothesis of these two statistics can be quite different.

Second, note that the distribution of $\max\{|\hat{\lambda}_{I,T}|\}$ is the distribution of the maximum of a sample of uncorrelated normal random variables, whereas the distribution of $\max\{|\hat{\lambda}_{L,T}|\}$ is the distribution of the maximum of a sample of correlated random variables, where the correlation depends on the model. For a stationary series the correlations can be very strong. For instance, if the series is white noise the estimation of $\widehat{w}_{\mathrm{LS}}^{(t)}$ is just the difference between the mean of the series before and after t, and the $\hat{\lambda}_{L,T}$ will be very correlated to $\hat{\lambda}_{L,T+1}$. On the other hand, if the series is a random walk then a LS is identical to an IO and the $\hat{\lambda}_{L,T}$ will be uncorrelated. This suggests that we should consider LS in a stationary time series as having a possible different effect than on a stationary time series.

In order to check the behavior of $\hat{\lambda}_{L,T}$ and $\hat{\lambda}_{I,T}$ in finite samples, we have carried out a large simulation study that we summarize here (the complete study is available from the author upon request). We present the results for three sample sizes, $n = 50, 100, 250$, and the 12 different models indicated in Table 1. Note that model 4, 5, and 6 are nonstationary whereas the others are stationary. For each combination of model and sample size 1000 time series replications were made. In each replication, the likelihood ratio statistics for IO and LS were computed at each observed point, and the maximum for each statistic was kept. The estimated residual standard deviation is computed using the omit-one method, that is, for testing a possible outlier at T, the residual standard deviation is calculated omitting the residual in this point. The simulations have been done using MATLAB (developed by The MathWorks Inc.). The a_t (random errors) are generated with $\sigma_a = 1$. The 95% percentile of the distribution of $\max\{|\hat{\lambda}_{i,t}|\}$, with $i = \mathrm{IO}, \mathrm{LS}$ is given in Table 2.

For stationary process, the main conclusions we draw from Table 2 are: (a) the critical values for testing for IO are similar for different models and we can safely use a value between 3.10 and 3.65 depending on the sample size; (b) the critical values for testing for a LS depend on the model, but they are always smaller than those for testing for an IO; (c) for models with $(1 - \sum_{j=1}^{i} \hat{\pi}_j)$ small for all i (models 2, 7, 8, and 10)

Table 1. Autoregressive parameter values.

Model	1	2	3	4	5	6	7	8	9	10	11	12
ϕ_1	0.2	0.8	−0.8	—	—	0.6	1	1.2	−0.2	1.15	−0.4	0.4
ϕ_2	—	—	—	—	—	—	−0.24	−0.4	−0.06	−0.36	−0.36	0.36
ϕ_3	—	—	—	—	—	—	—	—	—	0.105	0.144	−0.144
∇	—	—	—	1	2	1	—	—	—	—	—	—

Table 2. Critical values to 95% of the statistics $\max\{|\widehat{\lambda}_{I,t}|\}$ and $\max\{|\widehat{\lambda}_{L,t}|\}$ for the models in Table 1, where n is the sample size and M is the model number in Table 1, and recommended critical values.

n	M	IO	LS	M	IO	LS
50	1	3.15	2.52	7	3.10	2.89
50	2	3.09	2.89	8	3.09	3.05
50	3	3.10	2.39	9	3.11	2.36
50	4	3.14	3.34	10	3.12	3.14
50	5	3.16	3.30	11	3.10	2.22
50	6	3.16	3.40	12	3.16	2.63
100	1	3.39	2.61	7	3.36	2.93
100	2	3.36	3.03	8	3.39	3.19
100	3	3.35	2.41	9	3.40	2.50
100	4	3.41	3.61	10	3.41	3.25
100	5	3.40	3.48	11	3.36	2.52
100	6	3.43	3.55	12	3.37	2.70
250	1	3.62	2.72	7	3.71	3.10
250	2	3.73	3.22	8	3.64	3.13
250	3	3.67	2.59	9	3.66	2.59
250	4	3.60	3.73	10	3.66	3.28
250	5	3.69	3.75	11	3.62	2.59
250	6	3.69	3.74	12	3.66	2.90

Recommended

n	IO	LS-ST	LS-NST
50	3.10	2.60	3.35
100	3.35	2.75	3.55
250	3.65	2.90	3.75

ST = stationary models, NST = nonstationary models.

the difference between the critical values is small, as expected from the previous analysis, but for models with $(1-\sum_{j=1}^{i}\hat{\pi}_j)$ large (models 3, 9, and 11) this difference can be quite large.

For nonstationary models the critical values for LS are slightly larger than those for IO. From these results we conclude that an outlier detection method based on $\eta_t = \max_t\{|\hat{\lambda}_{i,t}|\}$ seems to be inadequate, because the sampling behavior of the maximum value of the statistic for the LS is different from the corresponding ones for IO and AO. This large difference will imply a very low power to detect level shift for stationary processes and a slight tendency to over detect level shift in nonstationary processes. The problem can be very serious because, as we will show in Sec. 5, an undetected level shift can have a strong effect on the parameter estimates. Thus, we recommend the critical values for $\hat{\lambda}_{I,T}$ and $\hat{\lambda}_{L,T}$ indicated at the bottom of Table 2.

4. UNMASKING SEQUENCES OF ADDITIVE OUTLIERS

When the time series contains several outliers, the contaminated model (2) can be generalized as

$$z_t = \sum_{i=1}^{k} \omega_i V_i(B) I_t^{(T_i)} + y_t \tag{5}$$

where k is the number of outliers. Assuming first that the parameters are known, and calling, as before, $e_t = \pi(B)z_t$, we have $e_t = x_t'\beta + a_t$ where $\beta' = (\omega_1, \dots, \omega_k)$ and $x_t' = (x_{1t}, \dots, x_{kt})$. It is well known that procedures that identify outliers one by one will work when the matrix $(\sum_{t=1}^{n} x_t x_t')^{-1}$ is roughly diagonal, but may lead to several biases when the series have patches of additive outliers and/or level shifts. Note that for an innovational outlier $x_{it} = I_t^{(T_i)}$, and therefore the estimation of its effect is typically uncorrelated with other effects. However, for additive outliers $x_{it} = \pi(B)I_t^{(T_i)}$ and the correlation between the effects of consecutive additive outliers can be very high. This is expected to happen when we have patches of outliers, an empirical fact found by Bruce and Martin (1989). For instance, suppose that in Eq. (5) we have $k = 2$ and two consecutive outliers at times T and $T + 1$ with magnitudes ω_1 and ω_2. Suppose that the parameters are known, and let $\hat{\omega}_1^{(*)}$ be the estimate of ω_1 when we use model (2) assuming that it is the only outlier. Then, the expected value of this estimate is given by

$$E(\hat{\omega}_1^{(*)}) = \omega_1 + \omega_2 \frac{\sum_{i=0}^{n-T-1} \pi_i \pi_{i+1}}{\sum_{i=0}^{n-T} \pi_i^2}$$

where $\pi_0 = -1$. As an example, if $\omega_1 = \omega_2 = \omega$ and the process is a random walk, the estimation assuming a single outlier at T will be $\omega/2$ and the variance of this estimate will be $\sigma_a^2/2$. Thus, the expected value for the likelihood ratio will be $\omega/\sqrt{2}\sigma_a$. On the other hand, in the correct model the expected value for the estimate is ω with variance $2\sigma_a^2/3$, leading to an expected likelihood ratio of $\sqrt{3}\omega/\sqrt{2}\sigma_a$. We see in this simple case that even if the parameters of the process are known, the presence of the second outlier makes the likelihood ratio test much smaller than the one from the correct model, making the identification of outliers in the sample more difficult. This is the masking effect, which is a very serious problem for outlier detection techniques.

A sequence of outliers is an even more serious problem when the parameters are unknown because then: (a) the estimate σ_a will be inflated, reducing the power of the detection tests; (b) the AR and MA parameter estimates can be strongly biased. To show this, suppose that we have a sequence of k additive outliers of sizes $\omega_1, \ldots, \omega_k$, at times $T, T+1, \ldots, T+k-1$. It is shown in the Appendix that calling the observed autocorrelation coefficients computed from the contaminated series $r_z(h)$, and the true autocorrelation coefficients without outlier effects $r_y(h)$, we have

$$r_z(h) = \frac{r_y(h) + n^{-1}S_1 + n^{-1}\left[\sum_{i=1}^{k-h} \tilde{\omega}_i\tilde{\omega}_{i+h} - \frac{k^2}{n}\tilde{\tilde{\omega}}^2\right]}{1 + n^{-1}\left[\sum \tilde{\omega}_i^2 - \frac{k^2}{n}\tilde{\tilde{\omega}}^2\right] + 2n^{-1}S_2} \tag{6}$$

where $\tilde{y}_t = (y_t - \bar{y})/s_y$, $\tilde{\omega} = \omega/s_y$, $\tilde{\tilde{\omega}} = \sum \tilde{\omega}_i/k$, $n{s_y}^2 = \sum y_t^2 - n\bar{y}^2$, $S_1 = \sum \tilde{\omega}_i(\tilde{y}_{T-1+i-h} + \tilde{y}_{T-1+h+i})$, and $S_2 = \sum \tilde{\omega}_i\tilde{y}_{T-1+i}$. We may consider the two following extreme cases. First, suppose that all the outliers have the same sign and similar size. Then, if $\omega \to \infty$, we have that $r_z(h) \to (k-h)/k$ for $k > h$, and $r_z(h) \to 0$ for $k \leq h$. This implies that if the number of outliers is small (k =1 or 2) the series will seem to be white noise, whereas if the number of outliers k is large the series will seem to be a non-stationary process. Second, suppose that the outliers are random values from a distribution of zero mean and very large variance. Then, it is easy to show from Eq. (6) that $r_z(h) \to 0$ and the series will seem to be white noise. In the general case of patches of outliers of arbitrary size, the effect depends on the relative size of each patch and the sizes and lengths of the other patches and can be obtained from Eq. (6).

Tsay et al. (2000) have shown that single multivariate outliers can produce patches of outliers in the univariate time series. Thus, we should be always aware of this possibility when searching for univariate outliers. The previous analysis suggests that: (a) in order to allow for the reduction of power due to masking, the step of initial outlier identification through the individual likelihood ratio test should be carried out with a moderate significance level (between 0.25 and 0.1), bearing in mind that the points will be checked jointly afterwards in the step of joint estimation.

(b) A step of search for patches of outliers should be included in the outlier identification procedure.

5. THE PROPOSED PROCEDURE

The procedure we propose for multiple outlier detection is based on the one developed by Chang et al. (1988), Tsay (1988), and Chen and Liu (1993) but includes several modifications to solve the problems indicated in the previous sections. It has three stages. In the first stage, *Initial parameter estimation*, a robust initial estimate is computed from a sample in which all influential points are eliminated. In the second stage, *Outlier detection*, outliers are identified one by one using the likelihood ratio test but the algorithm is modified to avoid the confusion between LS and IO. In the last stage, *Joint estimation*, the procedure uses maximum likelihood to jointly estimate the model parameters and the effects of the outliers and to search for patches of outliers.

5.1. Stage 1: Initial Estimation of the Model Parameters

The goal of this stage is to obtain a robust estimation of the model parameters. This is done by a two steps procedure. In the first step the parameters are estimated assuming no outliers and influential points in the computation of the parameters are found. In the second step these points are assumed to be missing values in the sample, and a new set of parameter estimates is computed from this modified time series data. The interpolation is easily carried out by introducing dummy variables at the points to be interpolated, as shown in Gómez et al. (1999). The parameter estimates obtained in the second step will be the initial parameter values used to search for outliers in the outlier detection procedure.

The intuitive basis of the method is to compute an initial robust estimate by cleaning the sample of all the influential points. Influential observations in time series can be classified as: (a) individually influential observations (e.g., a large additive outlier) and (b) jointly influential observations (e.g., a sequence of similar additive outliers). The first type of observations are detected by the statistic (Peña, 1991):

$$D_{\hat{Z}}(T) = \frac{(\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_T^{(\mathrm{INT})})'(\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_T^{(\mathrm{INT})})}{h\hat{\sigma}_a^2}, \tag{7}$$

where h is the number of ARMA parameters, $\hat{\sigma}_a^2$ is the estimate of the white-noise variance, $\hat{\mathbf{Z}}$ is the vector of forecasts assuming no outliers, and $\hat{\mathbf{Z}}_T^{(\mathrm{INT})}$ is the vector of forecasts computed by assuming that the T-th observation is an additive outlier. This vector of forecasts is obtained by fitting the intervention model $\pi(B)(z_t - \omega_A I_t^{(T)}) = a_t$, where $\pi(B)$, ω_A, and $I_t^{(T)}$ have been defined previously.

When the time series has a level shift, or a sequence of consecutive additive outliers of similar size, the points in the sequence can be jointly very influential, but they may not be individually influential observations. In fact, we have checked in a Monte Carlo study that the influence measure (7) detects a low percentage of the observations included in the sequence. Then, if we delete observations according only to $D_{\hat{Z}}(T)$, several outliers will be undetected, and they will bias the initial parameter estimates. A straightforward method to measure the effect of a level shift is

$$\mathrm{DL}(T) = \frac{(\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_T^{(\mathrm{ILS})})'(\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_T^{(\mathrm{ILS})})}{h\hat{\sigma}_a^2}, \tag{8}$$

where $\hat{\mathbf{Z}}$ is the vector of forecasts from the time series model without outlier effects, and $\hat{\mathbf{Z}}_T^{(\mathrm{ILS})}$ is the vector of forecasts from the intervention model $\pi(B)(z_t - \omega_L S_t^{(T)}) = a_t$, where ω_L is the effect of a LS at $t = T$ and $S_t^{(T)}$ is a step variable that takes the value 1 if $t \geq T$, and 0 otherwise.

Thus, the steps of stage 1 are as follows. (a) Calculate the estimates of the model parameters for the observed series, supposing that it is outlier free. (b) Calculate the influence measure $\mathrm{DL}(t)$ for every t, select the time at which the maximum value of $\mathrm{DL}(t)$ occurs, call it T_1, i.e., $T_1 = \arg\max \mathrm{DL}(t)$, then estimate the intervention model. If $\hat{\omega}_L$ is significant, remove the effect of the LS from the observations by defining the adjusted series

$$z_t^c = \begin{cases} z_t & t < T_1 \\ z_t - \hat{\omega}_L & t \geq T_1, \end{cases}.$$

On the adjusted series we compute again the influence measure $\mathrm{DL}(t)$ and repeat step 2 until $\hat{\omega}_L$ is non-significant. (c) Compute the individual measure of influence $D_{\hat{Z}}(t)$ for the adjusted series, z_t^c. Select the $\gamma\%$ more influential values and remove their effects from the observations as if they were additive outliers. Next, the model parameters are again estimated. These estimated parameters are used in stage 2.

5.2. Stage 2: Outlier Detection

The goal of this stage is to iteratively identify the presence of outliers in the time series. The steps are as follows. (a) Compute the residuals of the model using the final parameter estimates of stage 1, and $\hat{\sigma}_a$. (b) Compute $\hat{\lambda}_{I,t}$, $\hat{\lambda}_{A,t}$, and $\hat{\lambda}_{L,t}$, for $t = 1, \ldots, n$, and for each time t, let $\hat{\lambda}_{va,T_A}$ be the largest of statistics $|\hat{\lambda}_{i,t}|$ for IO and AO and T_A the corresponding time. If $\hat{\lambda}_{va,T_A} = |\hat{\lambda}_{I,T_A}| \geq C_1$, there is a possibility of IO at $t = T_A$. If $\hat{\lambda}_{va,T_A} = |\hat{\lambda}_{A,T_A}| \geq C_1$, there is a possibility of AO at $t = T_A$. (c) For $t = 1, \ldots, n$, select $\hat{\lambda}_{L,T_B} = \max_t |\hat{\lambda}_{L,t}|$. If $\hat{\lambda}_{L,T_B} \geq C_2$ there is a possibility of LS at $t = T_B$. There are three possible situations:

1. If neither outliers nor LS are found, then stop. The procedure finishes and the conclusion is that the observed series does not contain outliers.
2. An outlier (IO or AO) or a LS are detected. Then remove the effect of this outlier to obtain the adjusted series.
3. An outlier (IO or AO) and a level shift are detected. If they occur at different points, then remove both of them. If they occur at the same point, then a joint analysis is made to check if both effects are significant (it is possible that when there is a LS at T, the procedure detects at this time the LS, but simultaneously an IO or AO), by estimating an intervention model with two dummy variables at T: an impulse for IO/AO, and a step for the LS. The IO/AO and LS are considered as significant when the statistics $|\hat{\omega}_j / \widehat{dt(\hat{\omega}_j)}|$ are larger than the critical values C_1 and C_2 used in steps 3 and 4, where $\widehat{dt(\hat{\omega}_j)}$ is the estimated standard deviation of the estimated effect.

Steps (b) and (c) are iterated until the significant outliers are removed.

5.3. Stage 3: Joint Estimation

All the outliers identified in stage 2 and the model parameters are estimated jointly by an intervention model. If the effects of some outliers are not significant in this joint model, remove the smallest non-significant outlier from the set of detected outliers and estimate again the effect of the $k-1$ outliers. This process should be repeated until all the outliers in the final set are significant. Then, a search for patches of outliers is carried out as follows. If the number of time points between two consecutive additive outliers is g, and $g \leq h$, where h is chosen so that $|\pi_j| \leq 0.1$ for $j > h$, an intervention model for the adjusted series is estimated that includes impulse dummy variables for additive outliers for all the time points between those two consecutive outliers. In this way we check for sequence of outliers. If more outliers are found, a new adjusted series is defined and the procedure iterates through stages 2 and 3 until no more outliers are found.

5.4. Comments on the Procedure

In the first stage one possibility is to remove the effect of the individually influential values one by one, by estimating an additive outlier model for each of them. However, as the objective of this stage is to obtain initial parameter estimates, we prefer to delete the γ% more influential points. The value of γ depends on the number of expected outliers and, unless we are studying a very contaminated series, deleting 10% of the sample will be enough to obtain robustness in most applications. The search for outliers in stage 2 must be carried out with less strict critical values (e.g., $\alpha = 0.1$) than those used in the third stage, so that all the possible outliers are identified one by one. A consequence of this may be that this stage provides some wrong outliers, but this is not a serious problem since all the potential outliers detected will be afterwards jointly tested in stage 3. The values for C_1 and C_2 depend on the sample size and confidence level. For three sample sizes of 50, 100, and 250 observations we recommend the values of C_1 and C_2 that are shown in Table 2. Note that for stationary process the critical values for LS are smaller than for IO whereas for nonstationary models they are greater, in agreement with the simulation results shown in Table 2.

As we start by cleaning the time series for LS first one might think that, given the possible confusion between LS and IO, some IO would be identified as LS and these will produce an initial biased in the procedure. However note that the problem of confusion between LS and IO appears mainly because if we have a LS the estimated ARMA parameters are biased and the larger AR root will go to one (Chen and Tiao, 1990). However, the effects of IO on the parameters are small (Chang et al. 1988; Peña, 1990) and as the cleaning is carried out by searching for strong effects on the parameter values the possibility of confusion seems to be small. This expected result has been confirmed by a simulation study.

6. PERFORMANCE OF THE PROPOSED PROCEDURE

The performance of the procedure has been studied by simulation. As in Chen and Liu (1993) two measures of the performance of the procedure are considered: (a) the relative frequency of correct detection (type and location) of the outliers; (b) the accuracy and precision in the estimation of the model parameters. For this purpose, we use the sample mean and the sample root mean square error (RMSE) of the model parameter estimates. In the simulation study the factor levels have been selected in the same way as in Chen and Liu (1993) in order to facilitate the comparison between the two procedures. Five factors are considered: (a) type of outlier, IO, AO, and LS; (b) time series structure, AR(1) and MA(1); (c) outlier size, $3\sigma_a$, $4\sigma_a$, and $5\sigma_a$; (d) number of outliers, a single outlier or several adjacent outliers; (e) position of the outliers, at the beginning, in the middle or at the end of the series. The combination of these factors leads to the 15 models presented in Table 3. Models 1 through 12 correspond to the models used by Chen and Liu (1993). The simulations have been done using MATLAB and the random errors generated with $\sigma_a = 1$. The sample size is 100, the true values of the parameters ϕ_1 and θ_1 are both 0.6, and the number of replications is 500. The estimation of the variance ($\hat{\sigma}_a^2$) used to calculate the different statistics for testing outliers has been obtained by omitting the residual in $t = T$. The order of the model is supposed to be known.

Table 3. Models considered in the study of the performance of the proposed procedure.

Model AR (1)		Outliers		Model MA (1)		Outliers	
Model	Instants	Type	ω	Model	Instants	Type	ω
1	40	LS	3	4	40	LS	3
2	40	LS	4	5	40	LS	4
3	40	LS	5	6	40	LS	5
7	40, 41	AO, AO	3, 4	Model AR (1)		Outliers	
8	40, 41	IO, AO	3, 4	Model	Instants	Type	ω
9	40, 41	LS, AO	3, 4	13	50, 51, 52, 53	4 AO	5; −5; 5; −5
10	10, 11	AO, AO	4, −3	14	10, 11, 12, 13	4 AO	5; −5; 5; −5
11	10, 11	IO, AO	4, −3	15	88, 89, 90, 91	4 AO	5; −5; 5; −5
12	10, 11	LS, AO	4, −3				

Table 4. Comparison of the frequency of correct outlier detection, *D*, and false positives, *F*, with the procedure proposed in this article (PP) and the procedure of Chen and Liu (CL) for LS model.

		$\omega/\sigma_a = 3$		$\omega/\sigma_a = 4$		$\omega/\sigma_a = 5$	
		AR(1)	MA(1)	AR(1)	MA(1)	AR(1)	MA(1)
D	PP	0.35	0.68	0.71	0.89	0.91	0.93
D	CL	0.22	0.56	0.62	0.63	0.89	0.74
F	PP	0.21	0.40	0.10	0.39	0.29	0.44
F	CL	0.10	0.50	0.10	0.40	0.20	0.50

We do not present results about the detection of a single IO or AO, because the results for the proposed procedure are, as expected, almost identical to those given by Chen and Liu (1993) because then masking does not occur. We show in Table 4 the frequency of correct detection (type and location), D, and the average number of misidentified outliers or false positives, F, for the case of a single LS (models 1 to 6). The critical values used in our study correspond to $\alpha = 0.05$, $C_1 = 3.25$, and $C_2 = 2.75$. In order to make a fair comparison of our procedure with CL, we have used the same critical value as they do, $C = 3$. Table 5 presents the results of the parameter estimation of our procedure and, in brackets, those obtained by CL. In the table (*no*) corresponds to the estimated parameter supposing that there are no outliers, (*out*) to the last estimation of the parameter at the end of the procedure, and $\hat{\sigma}_a$ to the estimation of the residual standard deviation. RMSE($\hat{\theta}$) represents the sample root mean square error of the estimation of θ given by

$$\mathrm{RMSE}(\hat{\theta}) = \sqrt{(\mathrm{bias}(\hat{\theta}))^2 + \widehat{\mathrm{Var}}(\hat{\theta})}, \tag{9}$$

where $\mathrm{bias}(\hat{\theta}) = \hat{\theta} - \theta$, and $\widehat{\mathrm{Var}}(\hat{\theta})$ is the estimated sample variance of $\hat{\theta}$.

It is shown in Table 5 that in all the models for a single LS there is a clear improvement in the parameter estimation: the RMSE(*no*) is smaller with the proposed procedure and in model 6, this reduction is about 30%. The reduction of RMSE in the estimation of $\hat{\sigma}_a$ is also important and can be as large as 50%. For the case of two consecutive outliers, the comparison between the proposed procedure and the CL procedure shows that,

Table 5. Performance of the procedure for a single LS and for the models of two consecutive outliers. AR(1) and MA(1) models with $\phi = 0.6$ and $\theta = 0.6$. The results of the procedure of Chen and Liu (1993) are shown in brackets. M is the model number in Table 3. (*no*) represent the null case with no outliers and (*out*) with outliers.

M	RMSE (no)	RMSE (out)	RMSE ($\hat{\sigma}_a$)	M	RMSE (no)	RMSE (out)	RMSE ($\hat{\sigma}_a$)
1	0.214	0.140	0.117	7	0.090	0.094	0.088
	(0.213)	(0.189)	(0.234)		(0.087)	(0.112)	(0.192)
2	0.262	0.199	0.132	8	0.090	0.095	0.085
	(0.264)	(0.184)	(0.248)		(0.087)	(0.112)	(0.192)
3	0.292	0.162	0.137	9	0.184	0.143	0.184
	(0.296)	(0.166)	(0.196)		(0.183)	(0.212)	(0.216)
4	0.638	0.353	0.495	10	0.213	0.103	0.078
	(0.850)	(0.406)	(0.708)		(0.212)	(0.131)	(0.213)
5	0.692	0.329	0.746	11	0.147	0.010	0.077
	(0.962)	(0.511)	(1.238)		(0.212)	(0.131)	(0.213)
6	0.742	0.373	1.077	12	0.113	0.076	0.095
	(1.040)	(0.494)	(1.562)		(0.111)	(0.159)	(0.201)

again in all the models, the RMSE both for the parameters and $\hat{\sigma}_a$ is also smaller with our procedure. The critical values used in our procedure correspond to $C_1 = 3.25$, and $C_2 = 2.75$. For the CL procedure, the critical value is $C = 3$.

Table 6 shows the estimation results and the percentage of correctly detected outliers for models 13, 14, and 15 of Table 3. These models are not included in Chen and Liu (1993). The number of replications and the critical values are the same as before. In these three models the number of outliers is four. We see that although two or more outliers were correctly detected with a relative frequency in a range of 69–73.4%, less that 7% of the cases are the four outliers "correctly detected". In spite of this behavior, we have checked that the sample mean of the estimated autoregressive parameter after removing the detected outliers is quite close to the true value (0.6); the most unfavorable case is 0.561. Also, the RMSE(*out*), as seen in Table 6, is small, suggesting a good performance of the procedure. These apparently contradictory results, incorrect detection but correct estimation of the parameters, can be explained by the concept of equivalent configurations that we discuss in the next section.

Table 6. Performance of the procedure for multiple outliers and number of correctly detected outliers (D) with the proposed procedure for the models from 13 through 15 (percentage). (*no*) Represent the null case with no outliers and (*out*) with outliers.

				D				
Model	RMSE(no)	RMSE(out)	RMSE($\hat{\sigma}_a$)	0	1	2	3	4
13	0.567	0.096	0.150	1.80	29.2	38.2	24.6	6.2
14	0.575	0.100	0.178	1.20	25.4	41.0	26.4	6.0
15	0.571	0.123	0.227	1.00	26.8	40.8	24.6	6.8

6.1. Patches of Outliers and Equivalent Configurations

We say that two configurations of outliers are equivalent when their effects on a time series are indistinguishable. The equivalent configurations depend on the model. For instance, for an MA(1) model the configuration: two additive outliers at time T and $T+1$ of sizes ω_1 and ω_2 is equivalent to a configuration with two outliers at time T, the first AO of size $\omega_1 + \omega_2/\vartheta$ and the second IO of size $-\omega_2/\vartheta$.

When the parameters are estimated from the data, we can have a large number of approximately equivalent configurations. For instance, in models 13 through 15 we have an AR(1) with four consecutive additive outliers of similar size and alternating signs. Then, if the AR parameter ϕ is not large (so that $\phi^4 \simeq 0$), this sequence verifies:

$$\begin{aligned}&\omega I_t^{(T)} - \omega I_t^{(T+1)} + \omega I_t^{(T+2)} - \omega I_t^{(T+3)} \\ &\quad \simeq \omega(1-\phi B)^{-1} I_t^{(T)} - \omega_1 I_t^{(T+1)} + \omega_2 I_t^{(T+2)} - \omega_3 I_t^{(T+3)}\end{aligned}$$

and the sequence of outliers IO, AO, AO, AO with sizes $\omega_1 = \omega(1+\phi)$, $\omega_2 = \omega(1-\phi^2)$ and $\omega_3 = \omega(1+\phi^3)$ is a configuration (approximately) equivalent to the original one. As in our simulations $\phi = 0.6$ we expect that this last configuration can be found quite often in practice and this is confirmed in Table 7.

This table shows the configurations identified when two and three outliers of the sequences are correctly identified. For instance, in model 13 when three outliers are correctly identified, the sequence AO, IO, AO, AO, appears 46.34% of the time. When only two out of the four outliers are correctly identified, more possibilities may arise and Table 7 presents the more common configurations found in the simulations. For instance,

Table 7. Percentage of equivalent configurations for three and two correctly detected AO (models 13, 14, and 15).

Eq. Configurations	Model 13	Model 14	Model 15
Pattern detected when 3 outliers are correctly identified			
IO, AO, AO, AO	35.77	32.58	28.46
AO, IO, AO, AO	46.34	52.27	46.35
Others	17.89	15.15	25.19
Total	100	100	100
Pattern detected when 2 outliers are correctly identified			
IO, IO, AO, AO	11.00	9.76	12.25
IO, AO, IO, AO	8.90	10.73	2.94
AO, IO, IO, AO	2.62	0.05	0.49
IO, AO, −, AO	27.75	33.17	25.98
AO, IO, AO, −	12.57	11.70	7.84
Others	37.16	34.59	50.5
Total	100	100	100

as the third coefficient in the sequence IO, AO, AO, AO must be small compared to the others, it is possible that we would find no outlier in this position, leading to the sequence IO,AO,−,AO that is the most frequent in the three models compared in the table.

In summary, when dealing with a series that has consecutive outliers, we must be careful when comparing the correct detection performance of different procedures because of the lack of identification of multiple equivalent configurations.

6.2. Example

In this example, we use the variable Annual Unemployment Rate (person over 16 of the entire civilian labor force) from 1920 to 1979 of the set of time series provided by the SCA program. The CL procedure (implemented in the SCA program) identifies a TC in $t = 19$ with an ARI(1,1) model, and a critical value $C = 3$. Using the procedure proposed in this article, with an ARI(1,1) model, six outliers are detected: an AO at $t = 4$ and five IO at $t = 3, 11, 12, 13, 19$.

The proposed procedure applied to this series produces the following results: in stage 1, the maximum value of the influence measure for a LS, DL(t), occurs in t = 19. The effect of a LS in this point is non significant. The individual influence measure $D_{\hat{Z}}$ is calculated in every time point of the observed series. Table 8 shows the time points and the values of $D_{\hat{Z}}$ for 10% of the observations, those with the highest values of influence measure. These observations are then interpolated.

Stage 2 starts with the observed series and the estimation of the parameters obtained in stage 1. The procedure detects iteratively an IO in $t = 3$, an IO in $t = 19$, an IO in $t = 13$, an IO in $t = 11$, an IO in $t = 12$, and an AO in $t = 4$. These outliers are corrected iteratively, and the outlier detection starts again with the adjusted series and the estimated parameters obtained in stage 1. As there are no more outliers, stage 2 ends. In stage 3, the joint estimation is carried out and all the estimated effects and autoregressive parameters are significant. The final estimate parameters, their t values and the residual standard error are presented in Table 9.

As it is observed from Table 9, PP detects an IO at $t = 19$ and CL procedure (SCA) detects a TC at the same point. Both results could be considered as equivalent. However, the PP detects two significant outlier sequences that are not detected by the CL procedure.

Table 8. The individual measure of influence for the example.

t	4	5	10	13	18	19	39
$D_{\hat{Z}}$	2.4	1.7	1.5	0.9	3.1	2.8	0.9

Table 9. Parameter estimates for the ARI(1,1) model for the unemployment rate using the proposed procedure (PP) and the CL procedure.

PP			CL		
Type of parameter	Value	t Values	Type of parameter	Value	t Values
ϕ_1	0.71	6.70	ϕ_1	0.33	2.80
IO in $t=3$	−6.66	−4.7	TC in $t=19$	5.90	3.15
AO in $t=4$	−4.44	−5.2	σ^2	2.2367	
IO in $t=11$	4.24	2.6			
IO in $t=12$	5.55	3.1			
IO in $t=13$	6.15	4.1			
IO in $t=19$	5.85	4.7			
σ^2	1.6732				

7. CONCLUSION

In this article, we have presented a multiple outlier detection procedure that seems to improve the performance of previous methods in identifying level shifts and patches of outliers. This is achieved by using two tools. The first is a better initial parameter estimate that is obtained by cleaning the series of patches of jointly influential observations that are treated as level shifts. The second is a better significance level that avoids the confusion of level shift with innovative outliers. In our opinion, the proposed procedure keeps the powerful features of the original idea by Chang and Tiao (1983) and developed further by Tsay (1988) and Chen and Liu (1993), which have proved to be very useful in many applications, but adds some tools to prevent the confusion between IO and LS, to deal with multiple outliers and to avoid masking.

APPENDIX: THE EFFECT OF OUTLIERS ON THE AUTOCORRELATIONS COEFFICIENTS

We prove here Eq. (6). If we have the sequence of outliers $\omega_1, \ldots, \omega_k$ at time $T, \ldots, T+k-1$. Then we have

$$\begin{aligned}\sum z_t z_{t-h} - n\bar{z}^2 &= \sum y_t y_{t-h} - n\bar{y}^2 \\ &\quad + \sum \omega_i\big(y_{T-1+i-h} - \bar{y} + y_{T-1+h+i} - \bar{y}\big) + \sum_{i=1}^{k-h} \omega_i \omega_{i+h} - \frac{k^2}{n}\bar{\omega}^2.\end{aligned}$$

and

$$\sum z_t^2 - n\bar{z}^2 = \sum y_t^2 - n\bar{y}^2 + \sum \omega_i^2 - \frac{k^2}{n}\bar{\omega}^2 + 2\sum \omega_i(y_{T-1+i} - \bar{y}).$$

Calling $r_z(h) = (\sum z_t z_{t-h} - n\bar{z}^2)/(\sum z_t^2 - n\bar{z}^2)$ to the observed autocorrelation coefficient at lag h and $\tilde{y}_t = (y_t - \bar{y})/s_y$, $\tilde{\omega} = \omega/s_y$, $\tilde{\bar{\omega}} = \sum \tilde{\omega}_i$, $ns_y{}^2 = \sum y_t^2 - n\bar{y}^2$, $S_1 = \sum \tilde{\omega}_i(\tilde{y}_{T-1+i-h} + \tilde{y}_{T-1+h+i})$, $S_2 = \sum \tilde{\omega}_i \tilde{y}_{T-1+i}$ we have that

$$r_z(h) = \frac{r_y(h) + n^{-1}S_1 + n^{-1}\left[\sum_{i=1}^{k-h} \tilde{\omega}_i\tilde{\omega}_{i+h} - \frac{k^2}{n}\tilde{\bar{\omega}}^2\right]}{1 + n^{-1}\left[\sum \tilde{\omega}_i^2 - \frac{k^2}{n}\tilde{\bar{\omega}}^2\right] + 2n^{-1}S_2}.$$

ACKNOWLEDGMENT

This research has been supported by grant BEC2000-0167 from the MCYT, Spain.

REFERENCES

Balke, N. S. (1993). Detecting level shitfs in time series. *Journal of Bussiness and Economic Statistics* 11(1):81–92.

Bruce, A. G., Martin, D. (1989). Leave-*k*-out diagnostics for time series (with discussion). *Journal of the Royal Statistical Society B* 51(3):363–424.

Bustos, O. H., Yohai, V. J. (1986). Robust estimates for ARMA models. *Journal of the American Statistical Association* 81:155–168.

Chang, I., Tiao, G. C. (1983). Estimation of time series parameters in the presence of outliers. Technical Report 8, Statistics Research Center, University of Chicago.

Chang, I., Tiao, G. C., Chen, C. (1988). Estimation of time series parameters in the presence of outliers. *Technometrics* 30(2):193–204.

Chen, C., Liu, L. (1993). Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association* 88:284–297.

Chen, C., Tiao, G. C. (1990). Random level-shift time series models, ARIMA approximations, and level-shift detection. *Journal of the Business and Economic Statistics* 8(1):83–97.

Fox, A. J. (1972). Outliers in time series. *Journal of the Royal Statistical Society B* 34:350–363.

Gómez, V., Maravall, A., Peña, D. (1999). Missing observation in ARIMA models: skipping approach versus additive outlier approach. *Journal of Econometrics* 88:341–363.

Justel, A., Peña, D., Tsay, R. (2001). Detection of outlier patches in autoregresive time series. *Statistica Sinica* 11:651–673.

Le, N. H., Martin, R. D., Raftery, A. E. (1996). Modeling flat stretches, bursts, and outliers in time series using mixture transition distribution models. *Journal of the American Statistical Association* 91:1504–1515.

McCulloch, R. E., Tsay R. S. (1993). Bayesian inference and prediction for mean and variance shifts in autoregressive time series. *Journal of the American Statistical Association* 88:968–978.

McCulloch, R. E., Tsay R. S. (1994). Bayesian analysis of autoregressive time series via the Gibbs sampler. *Journal of Time Series Analysis* 15:235–250.

Peña, D. (1990). Influential observations in time series. *Journal of Business and Economic Statistics* 8:235–241.

Peña, D. (1991). Measuring influence in dynamic regression models. *Technometrics* (33):93–101.

Tsay, R. S. (1988). Outliers, level shifts, and variance changes in time series. *Journal of Forecasting* 7:1–20.

Tsay, R. S., Peña, D., Panktartz, A. (2000). Outliers in multivariate time series. *Biometrika* 87:789–804.