Working Paper 97-76
Statistics and Econometrics Series 27
October 1997

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624-9849

THE IDENTIFICATION OF MULTIPLE OUTLIERS IN ARIMA MODELS

María Jesús Sánchez and Daniel Peña*

Abstract

The presence of outliers causes biases in the estimation of ARIMA models. In this work we present a procedure for detecting outliers and obtaining a robust estimator of the parameters in univariate ARIMA time series models. There are three main problems in the existing procedures for detecting outliers in ARIMA time series models. The first one is the confusion between level shifts and innovative outliers when a level shift is present in a time series. The procedure includes a possible solution to avoid this problem based on not comparing the statistics for level shifts and innovative outliers together, because the critical values under the null hypothesis of no outliers can be quite different. The second problem is the biased estimation of the initial parameter values. In the existing procedures, this initial estimation is done under the hypotheses of no outliers in the data, which may lead to begin the search for outliers using a very biased set of parameters and, therefore, these procedures may fail. In order to solve this problem, we use two measures of influence in the first stage of the proposed procedure; one measure for individually influential observations, and an additional measure for level shifts and sequences of outliers. The third problem is masking. This problem appears when there is a sequence of additive outliers, because the usual one by one outlier identification method may fail in the identification of some of the members of the group. The proposed procedure seems to solve the aforementioned problems and obtains food parameter estimates when the time series has isolated outliers and/or multiple adjacent outliers. The performance of the proposed procedure is analyzed and an example is shown.

Key Words

Equivalent configurations; influential observations; misspecification; multiple outliers; robust estimation.

*Sánchez, Laboratorio de Estadística, Universidad Politécnica de Madrid; Peña, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, e-mail: dpena@est-econ.uc3m.es.

The Identification of Multiple Outliers in ARIMA Models

María Jesús Sánchez
Laboratorio de Estadística
Universidad Politécnica de Madrid

Daniel Peña
Dept. Estadística y Econometría
Universidad Carlos III de Madrid

July 18, 1997

Abstract

The presence of outliers causes biases in the estimation of ARIMA models. In this work we present a procedure for detecting outliers and obtaining a robust estimator of the parameters in univariate ARIMA time series models. There are three main problems in the existing procedures for detecting outliers in ARIMA time series models. The first one is the confusion between level shifts and innovative outliers when a level shift is present in a time series. The procedure includes a possible solution to avoid this problem based on not comparing the statistics for level shifts and innovative outliers together, because the critical values under the null hypothesis of no outliers can be quite different. The second problem is the biased estimation of the initial parameter values. In the existing procedures, this initial estimation is done under the hypotheses of no outliers in the data, which may lead to begin the search for outliers using a very biased set of parameters and, therefore, these procedures may fail. In order to solve this problem, we use two measures of influence in the first stage of the proposed procedure: one measure for individually influential observations, and an additional measure for level shifts and sequences of outliers. The third problem is masking. This problem appears when there is a sequence of additive outliers, because the usual one by one outlier identification method may fail in the identification of some of the members of the group. The proposed procedure seems to solve the aforementioned problems and obtains good parameter estimates when the time series has isolated outliers and/or multiple adjacent outliers. The performance of the proposed procedure is analyzed and an example is shown.

Keywords: Equivalent Configurations, Influential observations, Misspecification, Multiple Outliers, Robust Estimation.

1 Introduction

Time series data often have outliers or discordant observations. Identifying and managing these observations is necessary because they can produce pernicious effects in model specification. Even when the model is well specified, discordant observations can lead to wrong parameter estimation and, therefore, resulting forecasts can be expected to be poor.

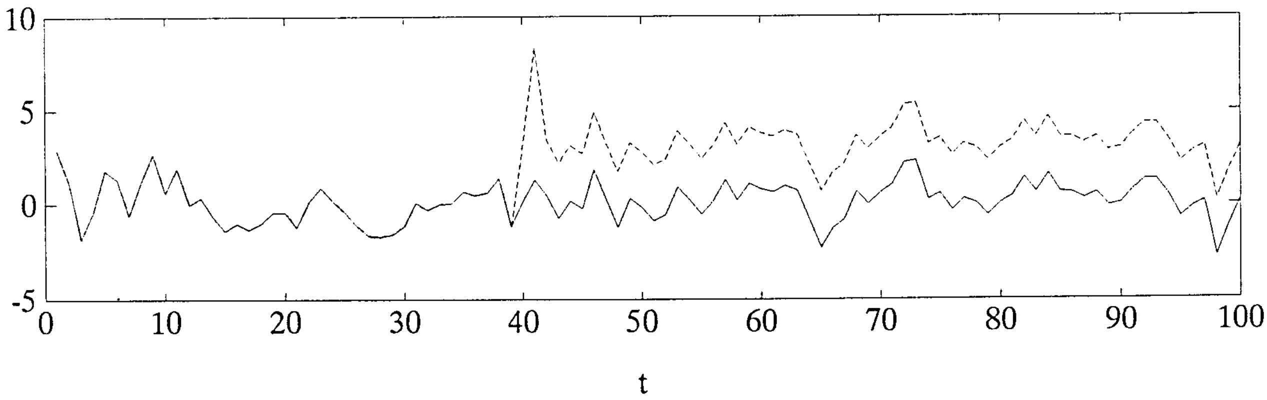
The study of outliers in time series can be carried out using different statisticals models that are proposed in the literature. In this paper, we deal with ARIMA models (Box and Jenkins, 1976), and more precisely in diagnostic methods.

Fox (1972) defines additive and innovative outliers and proposes the use of maximum likelihood ratio tests for detecting them. Chang and Tiao (1983) and Chang, Tiao and Chen (1988) extend the results of Fox (1972) to ARIMA models and present an iterative procedure for detecting these outliers and estimating the model parameters. Tsay (1988) generalizes this procedure for detecting Level Shifts and Temporary Changes. Balke (1993) proposes a modification to the Tsay Procedure for solving the confusion between Level Shift and Innovative outliers using an additional search of outliers with a white-noise model. This modification causes two problems: (1) the ARMA(0,0) model does not distinguish between innovative outliers and additive outliers, and (2) the white-noise search can identify spurious level shifts. Chen and Liu (1993) present an outlier detection and parameter estimation procedure that seems to be the most robust of all. However, this procedure may missidentify level shifts as innovative outliers and, besides, some outliers may not be identified due to masking effects. All of this can produce bias in parameter estimates.

An outlier is not necessarily an influential observation. Peña (1987, 1990) presents a statistic to measure influential outliers. Bruce and Martin (1989) define two diagnostics, DC and DV, for solving masking in time series when there are outlier patches. The DC diagnostic measures the change in the estimate of the ARIMA coefficients whereas DV measures the change in the estimated variance. If the size of the influential observations set is greater than 5 (i.e. level shift or variance change), the diagnostics can be wrong.

There are three main problems in the existing procedures for detecting outliers in ARIMA time series models. The first one is the confusion between level shift and innovative outliers (in favour of the latter) when a level shift is present in a time series. This situation appears in Tsay(1988), Balke (1993) and Chen and Liu(1993). The second is the biased estimation of the initial parameter values. This problem is due to the fact that the initial estimation of the parameters is made under the hypotheses of no outliers in the data, which may lead to begin the search for outliers by using a very biased set of parameters and, as a consequence of this, the procedure may fail. The third problem is masking. It appears when there is a sequence of additive outliers, because the usual procedures based on the identification of outliers one by one may fail in the identification of some of the members of the group. In this article, we present a procedure which seems to solve the aforementioned problems.

Figure 1 Plot of the simulated AR(1) series (-) and the contaminated (- -) series with a LS and an AO.



For instance, we have a simulated $n = 100$ observations of an AR(1) time series model with

$\hat{\phi} = 0.435$, and $\sigma_a = 1$. Then we have introduced two consecutive outlier effects: a LS of magnitude $\omega_L = 3\sigma_a = 3$ from $t = 40$ to $t = 100$, and an AO of magnitude $\omega_A = 4\sigma_a = 4$ at $t = 41$. Figure 1 shows the plot of both the simulated and observed series.

With the procedure proposed by Chen and Liu (1993), and using a critical value $C = 3$ only an AO at t=41 is detected, its estimated effect is $\hat{\omega} = 5.04$, and the $t-$value is 5.93, and the estimation of the autorregresive parameter at the end of the procedure is $\hat{\phi} = 0.908$ with a $t-$value of 14.16, and a residual standard error $\hat{s}_R = 1.147$. With our proposed procedure and using a critical value $C_{LS} = 3$ for LS and a critical value $C_{IO/AO} = 3.25$ for IO and AO, both the LS at t=40 and the AO at t=41 are detected; their estimated effects are respectively $\hat{\omega} = 3.261$ and $\hat{\omega} = 5.037$, and their t statistics are respectively 16.29 and 6.04. The estimation of the autoregressive parameter at the end of the procedure is $\hat{\phi} = 0.429$ with a $t-$value of 4.97, and the residual standard error $\hat{s}_R = 0.906$.

Another formulation initiated by Harvey (1981) is based on structural models using the state space representation (Aoki, 1987) with unobserved components. The parameters of these models can be estimated through the Kalman Filter. Harvey and Koopman (1992) discuss the detection of outliers in these models, while West (1981),Kitagawa (1987), and Peña and Guttman (1989) and Taplin (1993) work in the robust estimation of Kalman filter.

Within the classical robust methods, Denby and Martin (1979) present a generalization of the M estimators (called by them GM estimators) for first order autoregressive processes. Martin (1980) and Bustos (1982) carry out a generalization for autoregressive processes of order p. Martin (1981) and Lee and Martin (1982) present the extension for ARMA models, and Martin , Samarov and Vandaele (1983) for ARIMA models using a robust procedure of iterative estimation with an algorithm of robust filtering. Bustos and Yohai (1986) present two types of robust estimators for ARMA models.

Within the bayesian analysis of outliers in time series, McCulloch and Tsay (1993,1994) analyze ouliers, level shifts and variance change models in autoregressive processes.Le, Martin and Raftery (1996) use a mixture transition distribution model which can capture outliers, non-Gaussian features and nonlinear features, and Le, Raftery and Martin (1996) propose a method based on *robust Bayes factors.* for the order selection problem in autoregressive process with additive outliers.

This paper is organized as follows. In section 2 we present the model and the notation. In section 3 we analyze the confusion between innovational outliers and level shifts under H_0. The relationship between the statistics for testing the existence of an innovational outlier and a level shift and the sampling behavior of the maximum of the test statistics are analyzed. In section 4 we present a procedure to obtain an initial robust estimation of the model parameters. In section 5, we describe the proposed method to identify multiple outliers. In section 6, the performance of the proposed procedure is studied and an example of the procedure is shown.

2 Model and Notation

Let y_t be a stochastic process following an ARIMA model

$$\phi(B) \nabla^d y_t = \theta(B)a_t, \quad (1)$$

where B is the backshift operator such that $By_t = y_{t-1}$, $\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \ldots - \theta_q B^q$ are polynomials in B of degrees p and q, respectively, with roots outside the unit circle, $\nabla = 1 - B$ is the difference operator, $\nabla^d y_t$ is the stationary series, and a_t is a white-noise sequence of iid $N(0, \sigma_a^2)$ variables. In this paper, we will work with three types of outliers : innovational outliers (IO), additive outliers (AO), and level shifts (LS). For this purpose, the following model is considered

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

that is, instead of observing the ARIMA process y_t we observed the process z_t, that is contaminated by one outlier. In (2) y_t follows (1), ω_i is the outlier effect, $V_i(B) = 1/\pi(B)$ where $\pi(B) = (1 - \pi_1 B - \pi_2 B^2 - \ldots) = \theta^{-1}(B) \nabla^d \phi(B)$ for IO, $V_i(B) = 1$ for AO, and $V_i(B) = 1/(1-B)$ for LS; $I_t^{(T_i)}$ is an impulse variable that takes the value 1 if $t = T_i$ and 0 otherwise. Then calling $e_t = \pi(B) z_t$ equation (2) can be written as

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

where for innovational outliers $\omega_i = \omega_I$ and $x_t = I_t^{(T)}$, for additive outliers $\omega_i = \omega_A$ and $x_t = \pi(B) I_t^{(T)}$, and for level shifts $\omega_i = \omega_L$ and $x_t = \pi(B)(1-B)^{-1} I_t^{(T)}$. Assuming that the model parameters are known, the least squares estimators of ω_i, $\tilde{\omega}_i$, in (3) are $\sum e_t x_t / \sum x_t^2$, which leads to $\tilde{\omega}_A = \rho_A^2 \pi(F) e_T$, $\tilde{\omega}_I = e_T$ and $\tilde{\omega}_L = \rho_L^2 l(F) e_T$; where $\rho_A^2 = (1 + \pi_1^2 + \ldots + \pi_{n-T}^2)^{-1}$, F is the forward-shift operator such that $F e_t = e_{t+1}$, $\rho_L^2 = (1 + l_1^2 + \ldots + l_{n-T}^2)^{-1}$, and $l(B) = \pi(B)/(1-B)$. The variance of $\tilde{\omega}_i$ is $\sigma_a^2 (\sum x_t^2)^{-1}$ and therefore $Var(\tilde{\omega}_A) = \rho_A^2 \sigma_a^2$, $Var(\tilde{\omega}_I) = \sigma_a^2$ and $Var(\tilde{\omega}_L) = \rho_L^2 \sigma_a^2$.

For a single outlier case the following hypotheses are usually considered

$$\begin{array}{lll} H_0: & \omega_I = \omega_A = \omega_L = 0 & \text{(no outlier)} \\ H_I: & \omega_I \neq 0 & \text{(only IO)} \\ H_A: & \omega_A \neq 0 & \text{(only AO)} \\ H_L: & \omega_L \neq 0 & \text{(only LS)} \end{array}$$

The likelihood ratio test statistics for testing H_0 versus H_I, H_A and H_L are respectively $\lambda_{I,T} = \tilde{\omega}_I / \hat{\sigma}_a$, $\lambda_{A,T} = \tilde{\omega}_A / \rho_A \hat{\sigma}_a$ and $\lambda_{L,T} = \tilde{\omega}_L / \rho_L \hat{\sigma}_a$. Under the null hypothesis of no outliers, the $\lambda_{j,T}$ ($j = I, A, L$) 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, Tsay (1988) suggests calculating

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

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

When the time series contains several outliers the generalization of (2) is

$$z_t = \sum_{i=1}^{k} \omega_i V_i(B) I_t^{(T_i)} + y_t \tag{4}$$

where k is the number of outliers. Assuming as before that the parameters are known, and also calling $e_t = \pi(B) z_t$, we have

$$e_t = x_t' \beta + a_t \tag{5}$$

where $\beta' = (\omega_1, \ldots, \omega_k)$ and $x_t' = (x_{1t}, \ldots, x_{kt})$.

The outlier identification procedures which estimate the effects of the outliers one by one used model (2) instead of model (5). These procedures 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 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 we have $k = 2$ and two consecutive outliers of magnitudes ω_1 and ω_2 at times T and $T+1$. Then the expected value for the estimator of $\hat{\omega}_1^*$ using model (2), and assuming that it is the only outlier, 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 $t = T$ will be one half of the true outlier value. This will produce a masking effect and can lead to wrong outlier identification. This problem can be overcome by a step of joint estimation of all the outlier candidates detected by means of the individual model (2). It also suggests that 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.

3 The confusion between Innovative Outliers and Level Shifts under H_0

Suppose that y_t follows an ARIMA model. Then, the statistics for testing the existence of IO and LS in $t = T$ are:

$$(IO) \quad \hat{\lambda}_{I,T} = \frac{\hat{e}_T}{\hat{\sigma}_a},$$

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

where $\hat{l}_1 = -1 + \hat{\pi}_1$, $\hat{l}_2 = -1 + \hat{\pi}_1 + \hat{\pi}_2$, $\ldots$, $\hat{l}_{n-T} = -1 + \sum_{i=1}^{n-T} \hat{\pi}_i$.

Then the relationship between $\hat{\lambda}_{I,T}$ and $\hat{\lambda}_{L,T}$ is:

$$\hat{\lambda}_{L,T} = \frac{\hat{\lambda}_{I,T} + \frac{\sum_{i=1}^{n-T} (\hat{e}_{T+i}(1 - \sum_{j=1}^{i} \hat{\pi}_j))}{\hat{\sigma}_a}}{(1 + \sum_{i=1}^{n-T} (-1 + \sum_{j=1}^{i} \hat{\pi}_j)^2)^{1/2}}. \quad (6)$$

Note that for an AR(1) model when $\phi \to 1$ both statistics are equal because then both models are identical. For AR(p) models the closer to unity each of the elements of the sequence ϕ_1, $\phi_1 + \phi_2$, $\ldots$, $\sum_{i=1}^{p} \phi_i$ are, the nearer the LS critical values will be to the IO critical values. For an invertible ARMA model, under H_0 (no outliers), when $t = T$ is not close to the end of the series, for large n 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 for invertible ARMA models the statistics for level shift and innovative outliers should not be compared together, because the critical values under the null hypothesis can be quite different. In order to check this result in finite samples we have carried out a simulation study of the distribution of these two statistics. The objective of this study is to obtain the critical values for the statistics defined in section 2 that will allow us to decide whether or not there exist outliers under different circumstances. The response variables of this study are the 99, 95 and 90 % percentiles of $max_t\{\mid \hat{\lambda}_{i,t} \mid\}$, with i = IO, AO, LS and $t = 1, \ldots, n$. These percentiles will be called critical values C_i (i=IO, AO, LS.). The factors considered in this simulation study are the variables which define the tests statistics. These factors are:

1. Sample size. There are four sample sizes n = 30, 50, 100 and 150.
2. Type of outlier. We consider three types: IO, AO and LS.
3. The autoregressive process. We use 21 different models: seven AR(1) models, seven AR(2) models and seven AR(3) models. The parameter values for the AR(1), AR(2) and AR(3) processes are presented in table 1

The tests statistics for IO, AO and LS depend also on the estimated residual standard deviation, which is obtained using the omit-one method. Thus, if we are testing a possible outlier in $t = T$, the residual standard deviation is calculated omitting the residual in $t = T$. The simulations have been done using MATLAB (developed by The MathWorks, Inc.). The a_t (random errors) are generated with the Monte Carlo method, with $\sigma_a = 1$. For this case, the number of replications is 500, and for each replication $max \mid \hat{\lambda}_{i,t} \mid$ is selected with i =IO, AO, LS and $t = 1, \ldots, n$. The simulations have been developed without outliers. In the appendix A, we present the 95 percentiles of $max \mid \hat{\lambda}_{i,t} \mid$ for IO and LS (the 99 and 90 percentiles of $max \mid \hat{\lambda}_{i,t} \mid$ for IO and LS, and the percentiles for AO can be obtained from the authors), four sample sizes (30, 50, 100 y 150) and the 21 models that appear on table 1. The results show that the percentiles or critical values for IO are larger than the critical values for LS when using the same sample size and ϕ values. This is a general behavior except for model numbers 4, 8, 11 and 15 of table 1, in which both critical values are similar. Note that this result is expected because all these four models have the largest AR root close to one.

From the results of the study we conclude that the detection method based on $\eta_t = max_t\{\mid \hat{\lambda}_{i,t} \mid\}$, as given previously seems to be inadequate, because the sampling behavior of the maximum value of the statistic for the LS is different to the corresponding ones for IO and AO. In order to avoid the confusion between LS and IO when detecting outliers, we propose not to compare all the statistics $\mid \hat{\lambda}_{i,t} \mid$ together, but instead of this compare IO vs AO on the one hand, and on the other hand deal with LS alone. Thus, the maximum $\mid \hat{\lambda}_i \mid$ for IO-AO is compared with a critical value C_1, and the maximum $\mid \hat{\lambda}_{L,t} \mid$ is compared with a critical value C_2. If from these two comparisons we detect both an IO-AO and a LS in the same instant $t = T$, we will build an intervention model with two dummy variables, a step for the LS in $t = T$, and an impulse for IO-AO in $t = T$, and estimate both effects to study if both are significant. This method will be detailed in section 5.

Table 1. Autoregressive parameter values.

Model	Order	Parameters		
1	1	ϕ_1=0.2		
2	1	ϕ_1=0.4		
3	1	ϕ_1=0.6		
4	1	ϕ_1=0.8		
5	1	ϕ_1=-0.3		
6	1	ϕ_1=-0.5		
7	1	ϕ_1=-0.8		
8	2	ϕ_1=1;	ϕ_2= -0.24	
9	2	ϕ_1=0.5;	ϕ_2=0.2	
10	2	ϕ_1=0.5;	ϕ_2=-0.2	
11	2	ϕ_1=1.2;	ϕ_2=-0.4	
12	2	ϕ_1=0.6;	ϕ_2=-0.3	
13	2	ϕ_1=-0.2;	ϕ_2=-0.06	
14	2	ϕ_1=0.2;	ϕ_2=0.48	
15	3	ϕ_1=1.15;	ϕ_2=-0.36;	ϕ_3=0.105
16	3	ϕ_1=-0.2;	ϕ_2=0.56;	ϕ_3=0.192
17	3	ϕ_1=0.2;	ϕ_2=0.56;	ϕ_3=-0.192
18	3	ϕ_1=1.15;	ϕ_2=-0.36;	ϕ_3=-0.105
19	3	ϕ_1=0.4;	ϕ_2=0.36;	ϕ_3=-0.144
20	3	ϕ_1=-1.1;	ϕ_2=0.17;	ϕ_3=0.315
21	3	ϕ_1=-0.4;	ϕ_2=-0.36;	ϕ_3=0.144

4 Measures of Influential Observations in ARIMA models

In time series, the estimation of model parameters can turn out wrong when there are outliers, and searching for outliers using these biased parameter estimates can lead to incorrect results. In order to start the searching procedure with better parameter estimates we propose to compute them from a sample in which all data points that have large influence on the parameter computation are assumed to be missing values. Thus we compute an initial robust estimate by cleaning the sample from all the influential points, which are substituted by their interpolated values using the rest of the sample, so that the resulting estimated parameters are free of their effects.

Influential observations in time series can be classified in two types: 1-. isolated observations, that we shall call *individually influential observations* (i.e. an additive outlier) and 2-. sets of observations, that we shall call *jointly influential observations* (i.e., LS or outlier sequence). Peña (1991) suggests to measure the influence of a single observation using the statistic $D_{\hat{Z}}(T)$

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

where h is the order of the ARIMA model or number of parameters, $\hat{\sigma}_a^2$ is the estimation of the white-noise variance, $\hat{\mathbf{Z}}$ is the vector of forecasts assuming no outliers, and $\hat{\mathbf{Z}}_T^{(INT)}$ is the vector of forecasts computed by assuming that the $T-th$ observation is an additive outlier. This vector of forecasts is provided by the intervention model

$$\pi(B)(z_t - \omega_A I_t^{(T)}) = a_t, \tag{8}$$

where $\pi(B)$, ω_A and $I_t^{(T)}$ have been defined previously.

A different way to measure the influence of a single observation is by analyzing the change it produces on the parameter estimates. Let $\hat{\boldsymbol{\pi}}$ be the maximum likelihood estimate (MLE) of $\boldsymbol{\pi}$ supposing than there are not outliers, and let $\hat{\boldsymbol{\pi}}_{(T)}$ be the MLE considering that the observation in $t = T$ is missing. The measure of influence, $P_\pi(T)$, Peña (1990) is the Mahalanobis distance between the vectors $\hat{\boldsymbol{\pi}}$ y $\hat{\boldsymbol{\pi}}_{(T)}$.

$$P_\pi(T) = \frac{(\hat{\boldsymbol{\pi}} - \hat{\boldsymbol{\pi}}_{(T)})' \hat{\Sigma}_\pi^{-1} (\hat{\boldsymbol{\pi}} - \hat{\boldsymbol{\pi}}_{(T)})}{h\hat{\sigma}_a^2}, \tag{9}$$

where $\hat{\sum}_\pi \sigma_a^2$ is the variance-covariance matrix of the $\hat{\boldsymbol{\pi}}$ estimated vector, and h is the number of parameters. If $\hat{\mathbf{Z}} = \mathbf{X}_z \hat{\boldsymbol{\pi}}$ is the estimated vector of forecasts and $\hat{\mathbf{Z}}_{(T)} = \mathbf{X}_z \hat{\boldsymbol{\pi}}_{(T)}$ is the estimated vector of forecasts assuming that the observation in $t = T$ is missing, where $\mathbf{X}_z$ is the matrix

$$\mathbf{X}_z = \begin{pmatrix} z_h & z_{h-1} & \cdots & z_1 \\ z_{h+1} & z_h & \cdots & z_2 \\ . & . & \cdots & . \\ . & . & \cdots & . \\ . & . & \cdots & . \\ z_{t-1} & z_{t-2} & \cdots & z_{t-h} \end{pmatrix},$$

the influence measure (9) can be written as

$$P_{\hat{Z}}(T) = \frac{(\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_{(T)})'(\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_{(T)})}{h\hat{\sigma}_a^2}, \tag{10}$$

since $(\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_{(T)})'(\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_{(T)}) = (\hat{\boldsymbol{\pi}} - \hat{\boldsymbol{\pi}}_{(T)})' \hat{\Sigma}_\pi^{-1} (\hat{\boldsymbol{\pi}} - \hat{\boldsymbol{\pi}}_{(T)}) = (\hat{\boldsymbol{\pi}} - \hat{\boldsymbol{\pi}}_{(T)})' (\mathbf{X}_z' \mathbf{X}_z)(\hat{\boldsymbol{\pi}} - \hat{\boldsymbol{\pi}}_{(T)})$.

Note that in linear regression model a measure of influence based on the change in the parameter estimates is always equivalent to the one based on the change in forecast, whereas in time series this equivalence is lost, because (7) and (10) are different. In fact, it is proved in the appendix B that the relationship between these two statistics is:

$$D_{\hat{Z}}(T) \simeq P_{\hat{Z}}(T) + \frac{\hat{\lambda}_{A,T}^2}{h}, \tag{11}$$

so that $D_{\hat{Z}}(T)$ can be interpreted as the effect of the change of the parameters plus the effect of the additive outlier. We have found in a simulation study that statistic (7), which includes the outlier size is more effective to detect outliers which have a strong influence on the model than (10), and we will use this measure from now on in this paper.

When the time series has a LS or a sequence of consecutive additive outliers of a similar size, which produces a behavior similar to a LS, we have observed that the influence measure (7) detects as influential observations a low percentage of the observations affected by the LS or included in that sequence. This is the masking effect. Then, if we delete observations according only to $D_{\hat{Z}}(T)$ several outliers will be undetected, and will biased the initial parameter estimates. To avoid this situation we now develop a influence measure for a LS which will be used in the procedure jointly with (7) to carry out the initial cleanning of the sample data.

Let y_t follow an $ARIMA(p, d, q)$ process and let us assume that there is a LS in $t = T$ being z_t the observed series. Let $\hat{\boldsymbol{\pi}}$ and $\hat{\boldsymbol{\pi}}_L$ be the conditional maximum likelihood estimates of $\boldsymbol{\pi}$; where $\hat{\boldsymbol{\pi}}$ is obtained by assuming no outliers, whereas $\hat{\boldsymbol{\pi}}_L$ is computed assuming a LS in $t = T$ and using the model

$$\pi(B)(z_t - \omega_L S_t^{(T)}) = a_t, \tag{12}$$

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. We propose measuring the influence of a LS by considering the change in the forecast vector as

$$DL(T) = \frac{(\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_T^{(ILS)})'(\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_T^{(ILS)})}{h\hat{\sigma}_a^2}, \tag{13}$$

where $\hat{\mathbf{Z}}$ is the vector of forecasts assuming no LS, and $\hat{\mathbf{Z}}_T^{(ILS)}$ is the vector of forecasts from the intervention model (12). As in the case of additive outliers, we could have measure the influence of LS by analyzing the change in the parameter estimates. Calling, $\hat{\boldsymbol{\pi}}$ and $\hat{\boldsymbol{\pi}}_L$ to the MLE of $\boldsymbol{\pi}$; $\hat{\boldsymbol{\pi}}$ supposing that there are no outliers, and $\hat{\boldsymbol{\pi}}_L$ considering that there exists a LS in $t = T$ and using the model (12), we could use the formulation (9) to build a measure of the change in the parameters.

We obtained, in the appendix C, the measure of the change in the parameters for a LS, $PL(T)$, and it is proved that the relationship between the statistics $DL(T)$ and $PL(T)$ is:

$$DL(T) \simeq PL(T) + \frac{\hat{\lambda}_{L,T}^2}{h}, \tag{14}$$

and, as in the additive outlier case, $DL(T)$ can be interpreted as the effect of the change of the parameters plus the effect of a level shift.

In summary, for detecting individually and jointly influential observations we propose to use two measures of influence: $D_{\hat{Z}}(T)$ for individually influential observations, and $DL(T)$ for LS and sequences of outliers.

5 The Proposed Procedure

The procedure we propose for multiple outlier detection has three stages. In the first stage, *Initial parameter estimation*, a robust initial estimate is computed from a sample in which all influential points either individually (as measured by (7)) or jointly (as measured by (13)) are eliminated. In the second stage, *Outlier detection*, outliers are identified using a similar algorithm to the one developed by Chang, Tiao and Chen (1988), Tsay (1988) and Chen and Liu (1993), 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. This stage finishes with a new outlier detection step using the maximum likelihood parameter estimates.

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. Thus, the procedure uses the measures of influence (7) and (13), to identify the observations that can produce bias estimates.

The steps of stage 1 are:

1. Calculate the estimates of the model parameters using the observed series and supposing that there are no outliers on the time series.

2. Calculate the influence measure $DL(t)$ for every t. Select the time at which the maximum value of this $DL(t)$ occurs and call it T_1, i.e. $T_1 = \arg\max DL(t)$. Then estimate the intervention model

$$\pi(B)(z_t - \omega_L S_t^{(T_i)}) = a_t$$

where z_t is the observed series, $\pi(B)$ is the autoregressive approximation of the model, $S_t^{(T_i)}$ is a step variable at $t = T_i$ with $i = 1$, and ω_L is the effect of the LS at time T_i.

   (a) If $\hat{\omega}_L$ is significant, remove the effect of the LS from the observations by defining the adjusted series

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

   and calculate again the influence measure $DL(t)$ and repeat step 2. The process is repeated until $\hat{\omega}_L$ is not significant.

   (b) If $\hat{\omega}_L$ is not significant, go to step 3.

3. With the last adjusted series, z_t^c, the individual measure of influence $D_{\hat{Z}}(t)$ is calculated. Select the $\alpha\%$ more influential values, 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.

The output of this stage is a set of robust parameter estimates, once the effects of significant LS and an $\alpha\%$ of individually influential observations have been removed from the sample.

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:

1. Compute $\hat{\sigma}_a$ and the residuals of the model using the parameters of stage 1.

2. Compute $\hat{\lambda}_{I,t}$, $\hat{\lambda}_{A,t}$ and $\hat{\lambda}_{L,t}$, for $t = 1, \ldots, n$.

3. For each time t, the statistics $|\hat{\lambda}_{i,t}|$ for IO and AO are compared. Let $\hat{\lambda}_{va,t}$ be the largest of them and $\hat{\lambda}_{va,T_A}$ the maximum value of $\hat{\lambda}_{va,t}$, which occurs at time $T = T_A$. If $\hat{\lambda}_{va,T_A} = |\hat{\lambda}_{I,T_A}| \geq C_1$, there is a possibility of IO at $t = T_A$. The C_1 value is a predeterminate critical value for IO and AO, and it depends on the sample size, model structure and confidence level. If $\hat{\lambda}_{va,T_A} = |\hat{\lambda}_{A,T_A}| \geq C_1$, there is a possibility of AO at $t = T_A$.

4. For $t = 1, \ldots, n$, select $\hat{\lambda}_{L,T_B} = max_t \mid \hat{\lambda}_{L,t} \mid$. If $\hat{\lambda}_{L,T_B} \geq C_2$ there is a possibility of LS at $t = T_B$. The C_2 value is a predeterminate critical value for LS, and it depends on the sample size, model structure and confidence level.

5. There are four possible situations:

   (a) If neither outliers nor LS are found, then stop. The procedure finishes and the conclusion is that the observed series is free from outlier effects.

   (b) Innovative or additive outliers are detected, but no LS. By removing the effect of the IO or the AO, the *adjusted series* is obtained.

   (c) A LS is detected, but no innovative or additive outliers. By removing the effect of the LS, the *adjusted series* is obtained.

   (d) Innovative or additive outliers and level shifts are detected. If these occur in different points, then remove both of them. If these outliers occur in the same point, then study if their effects are both significant (it is possible that when there is a LS in $t = T$, the procedure detects at this time the LS but simultaneously an IO or AO). For this purpose, the procedure uses an intervention model with two dummy variables: an impulse in $t = T$ for IO/AO, and a step in $t = T$ for the LS. The IO-AO and LS in $t = T$ are significant when the statistics $\mid \hat{\omega}_j / \widehat{dt(\hat{\omega}_j)} \mid$ are larger than the different critical values used in steps 3 and 4 (C_1 for IO/AO, and C_2 for LS). $\widehat{dt(\hat{\omega}_j)}$ is the estimated standard deviation of the estimate effect for the outlier in $t = T$.

   Step 5 finishes when all the significant outliers are removed.

6. Using the *adjusted series* (free from the detected IO or AO and/or LS effects), and using the parameter estimation in stage 1, go to (2) and iterate through (2), (3), (4) y (5) until no additional outliers are detected.

The inputs of stage 2 are the estimation of the parameters at stage 1 and the observed series. The outputs are the time points, type and estimated effects for all the detected outliers.

5.3 Stage 3: Joint estimation.

Calculate the residuals of the model ($\hat{e}_t$) by filtering the observed series using the estimated parameters after correcting the k outliers detected. The outlier effects are jointly estimated using

$$\hat{e}_t = \sum_{i=1}^{k} \omega_i \hat{\phi}(B) V_i(B) I_t^{(T_i)} + a_t \tag{15}$$

where $V_i(B) = 1/\hat{\phi}(B)$ for an IO, $V_i(B) = 1$ for an AO, and $V_i(B) = 1/(1-B)$ for a LS, k is the number of detected outliers (or the number of interventions of the model) which occur at time points $T_1, \ldots, T_k$.

Test if the effects of the outliers are significant. If the effects of some outliers are not significant, 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 obtain the adjusted series by removing their effects from the observed series. With the last series, the model parameters are estimated, and then iterate through stages 2 and 3. In these two stages, the procedure uses the adjusted series and the last parameter estimates. When no more outliers are found in stage 2, remove all the effects of the detected outliers from the observed series obtaining *the adjusted series*. Finally, identify the model for the adjusted series, and jointly estimate the outlier effects and the model parameters.

Comments to the procedure: The search of the different outliers in stage 2 must be carried out with less strict critical values (i.e. $\alpha = 0.1$) than those used in the third stage. Thus, all the possible outliers are identified one by one. It is possible that this stage provides some wrong outliers, but this is not a problem since all the outliers are afterwards jointly estimated in (15).

6 Performance of the Proposed Procedure

In order to study the performance of the procedure, we carried out a simulation study. The evaluation criterion of the procedure is similar to the one of Chen and Liu (1993), CL from now on. Two measures of the procedure performance are considered:

1. The relative frequency of correct detection (type and location) of the outliers.
2. 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 Chen and Liu (1993) in order to facilate the comparison between the procedures. The factors are those that can influence the outlier detection and the estimation of model parameters. These factors are: 1) Type of outlier: IO, AO and LS; 2) Time series structure: AR(1) and MA(1); 3) Size of outlier: $\omega = 3\sigma_a = 3$, $\omega = 4\sigma_a = 4$ and $\omega = 5\sigma_a = 5$; 4) Number of outliers: a single outlier or several adjacent outliers; 5) 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 29 cases presented in Table 2. Cases 1 to 24 are replications of ARMA models that appear in Chen and Liu (1993).

The simulations have been done using MATLAB (developed by The MathWorks, Inc.). The random errors are generated with the Monte Carlo method, with $\sigma_a = 1$. The sample size of the time series is 100. The true values of the parameters ϕ_1 and θ_1 are both 0.6 and $\sigma_a^2 = 1$. For every case of table 2. 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 with the *omit-one method*. Thus, if we test a possible outlier at $t = T$, the residual standard deviation is obtained by omitting the residual in $t = T$. Once all the detected outliers are corrected, $\hat{\sigma}_a$ is the residual standard deviation. The order of the model is supposed to be known. The estimation of the autoregressive parameter has been obtained by least squares, and this for the moving average parameter by the Galbraith and Zinde-Walsh (1994) method, with a third order autoregressive approximation. For instance, if $\hat{\phi}_1$, $\hat{\phi}_2$ and $\hat{\phi}_3$ are the estimated parameters, the estimation of the moving-average parameter is $\hat{\theta} = -\hat{\phi}_1$.

Table 2. Different cases for studying the performance of the proposed procedure.

CASE	Model	Outliers
1	AR(1)	AO t=40 $\omega = 3$
2	AR(1)	AO t=40 $\omega = 4$
3	AR(1)	AO t=40 $\omega = 5$
4	AR(1)	IO t=40 $\omega = 3$
5	AR(1)	IO t=40 $\omega = 4$
6	AR(1)	IO t=40 $\omega = 5$
7	AR(1)	LS t=40 $\omega = 3$
8	AR(1)	LS t=40 $\omega = 4$
9	AR(1)	LS t=40 $\omega = 5$
10	AR(1)	AO t=40 $\omega = 3$ AO t=41 $\omega = 4$
11	AR(1)	IO t=40 $\omega = 3$ AO t=41 $\omega = 4$
12	AR(1)	LS t=40 $\omega = 3$ AO t=41 $\omega = 4$
13	AR(1)	AO t=10 $\omega = 4$ AO t=11 $\omega = -3$
14	AR(1)	IO t=10 $\omega = 4$ AO t=11 $\omega = -3$
15	AR(1)	LS t=10 $\omega = 4$ AO t=11 $\omega = -3$
16	MA(1)	AO t=40 $\omega = 3$
17	MA(1)	AO t=40 $\omega = 4$
18	MA(1)	AO t=40 $\omega = 5$
19	MA(1)	IO t=40 $\omega = 3$
20	MA(1)	IO t=40 $\omega = 4$
21	MA(1)	IO t=40 $\omega = 5$
22	MA(1)	LS t=40 $\omega = 3$
23	MA(1)	LS t=40 $\omega = 4$
24	MA(1)	LS t=40 $\omega = 5$
25	AR(1)	4 AO t=50,51,52,53 $\omega = 5; -5; 5; -5$
26	AR(1)	4 AO t=75,76,77,78 $\omega = 5; -4; 5; -4$
27	AR(1)	4 AO t=50,51,52,53 $\omega = 5; -3; 5; -3$
28	AR(1)	4 AO t=10,11,12,13 $\omega = 5; -5; 5; -5$
29	AR(1)	4 AO t=10,11,12,13 $\omega = 5; -3; 5; -3$

In the detection of a single IO or AO the results for the proposed procedure are almost identical to those given by Chen and Liu (1993). This is expected, because then masking does not occur. For this reason, table 3 only shows the frequency of correct detection (type and location) for the cases of a single LS. The selection of the critical values is as follows (see appendix A):

1. In the AR(1) models, $C_1 = C_{AI} = 3.25$ is the critical value for the detection of IO and AO with $\alpha = 0.05$, and $C_2 = C_{LS} = 2.85$ is the critical value for the detection of the LS with $\alpha = 0.05$. These values are used in all the AR(1) cases, except in cases 7,8 and 9, where the time series has a single LS. We use $C_{LS} = 3$ for comparing the results of our procedure against the results of the procedure of Chen and Liu (CL).

2. In the MA(1) models, we use $C_1 = C_{AI} = 3.3$ for detecting IO and AO, and $C_2 = C_{LS} = 2.75$ for detecting LS, these values have been obtained with $\alpha = 0.05$.

In order to make the comparison of the proposed procedure versus the CL procedure we have chosen the critical value corresponding to the existing outlier. Thus, in the proposed procedure two critical values are used, one for AO and IO and another for LS. When we want to compare the results of both procedures for a single outlier, the results of the CL procedure correspond to the existing outlier. For example, if a single IO is presented, then $C = 3.25$ in the CL procedure; but if the presented outlier is a LS and the model is MA(1), the selected critical value is $C = 2.75$.

Table 3. Comparison of the procedure proposed in this paper (PP) and the procedure of Chen and Liu (CL) for LS model.

	$\omega = 3\sigma_a = 3$		$\omega = 4\sigma_a = 4$		$\omega = 5\sigma_a = 5$	
	AR(1)	MA(1)	AR(1)	MA(1)	AR(1)	MA(1)
PP	0.31	0.70	0.69	0.90	0.89	0.93
CL	0.22	0.51	0.62	0.65	0.89	0.75

Tables 4 and 5 present the results of the parameter estimation. Those obtained by Chen and Liu (1993) appear between brackets. PAR is the sample mean of the estimated parameter, supposing that there are no outliers; PARF is the sample mean of the last estimation of the parameter at the end of the procedure, $\hat{\sigma}_a$ is the sample mean of the estimation of the residual standard deviation, and RMSE($\hat{\theta}$) represents the sample root mean square errors of the estimation of the corresponding parameter, which is calculated using the following equation:

$$\mathrm{RMSE}(\hat{\theta}) = \sqrt{(\mathrm{bias}(\hat{\theta}))^2 + \widehat{\mathrm{Var}(\hat{\theta})}} \qquad (16)$$

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

Again we only present in table 4 the results of parameter estimation for a single LS. For the cases of a single outlier, the results are similar, as expected, with both procedures. For LS, the sample mean of the estimated autoregressive or moving average parameter (PARF) is closer to the true value ($\phi = 0.6$ or $\theta = 0.6$) with the proposed procedure than with the CL procedure in four out of the six cases. In all these cases the RMSE is smaller with the proposed procedure. In five out of the six cases, the sample mean of $\hat{\sigma}_a$ obtained with the proposed procedure is closer to the true value ($\sigma_a = 1$) than the sample mean obtained with the CL procedure . In all these cases, the RMSE is smaller.

In table 4, the sample mean of the estimated parameter θ (PAR) in the cases 22, 23 and 24 is closer to the real value because in these cases the Galbraith and Zinde-Walsh (1994) method is used. This approximation is specially good when there are outliers, as in these cases.

Table 5 presents the results of the parameter estimation for two adjacent outliers (cases 10 through 15). The critical value for IO and AO is $C_{AI} = 3.25$ with $\alpha = 0.05$, and the critical value for LS is $C_{LS} = 2.85$ with $\alpha = 0.05$. In the results of the CL procedure for cases 12 and 15 three different possibilities are presented: these are $C = 3.25$, 3.0 and 2.75 respectively. The reason for these three comparisons is that one LS and one AO are studied, and in the proposed procedure we compare the IO and AO with a critical value ($C_{AI} = 3.25$) and the LS with another ($C_{LS} = 2.85$), while the CL procedure uses a single critical value for all outliers (IO,AO and LS).

Table 4 Performance of the procedure for a single LS. AR(1) and MA(1) models with $\phi = 0.6$ and $\theta = 0.6$. C_{AI} is the critical value for identifying IO and AO, and C_{LS} is the critical value for the detection of LS. The results of the procedure of Chen and Liu (1993) are shown between brackets.

Case	C_{AI}	C_{LS}	PAR	RMSE(PAR)	PARF	RMSE(PARF)	$\hat{\sigma}_a$	RMSE($\hat{\sigma}_a$)
7	3.25	3.00	0.809	0.214	0.720	0.144	1.071	0.115
	(3.00)	(3.00)	(0.808)	(0.213)	(0.763)	(0.189)	(1.124)	(0.234)
8	3.25	3.00	0.860	0.262	0.770	0.206	1.081	0.129
	(3.00)	(3.00)	(0.862)	(0.264)	(0.723)	(0.184)	(1.094)	(0.248)
9	3.25	3.00	0.894	0.542	0.703	0.163	1.094	0.136
	(3.00)	(3.00)	(0.895)	(0.296)	(0.692)	(0.166)	(1.036)	(0.196)
22	3.3	2.75	-0.029	0.635	0.387	0.316	1.312	0.459
	(2.75)	(2.75)	(-0.248)	(0.850)	(0.315)	(0.443)	(1.610)	(0.993)
23	3.3	2.75	-0.084	0.689	0.410	0.328	1.424	0.745
	(2.75)	(2.75)	(-0.360)	(0.962)	(0.268)	(0.467)	(1.829)	(1.287)
24	3.3	2.75	-0.134	0.739	0.429	0.357	1.598	1.121
	(2.75)	(2.75)	(-0.438)	(1.040)	(0.257)	(0.463)	(1.940)	(1.497)

For the cases of two adjacent outliers, the comparison between the proposed procedure and the CL procedure when estimating the parameters after removing the detected outliers, shows that in 66.7% of the cases, the sample mean of the estimated autoregressive parameter is closer to the true value ($\phi = 0.6$) with the procedure proposed, and in all these cases, the RMSE is also smaller. In 67% of the cases, the sample mean of $\hat{\sigma}_a$ obtained with the proposed procedure is closer to the true value ($\sigma_a = 1$), and in all these cases, the RMSE is smaller with the proposed procedure than with the CL procedure.

Table 5 Performance of the procedure for the cases of two adjacent outliers. AR(1) and MA(1) models with $\phi = 0.6$ and $\theta = 0.6$. C_{AI} is the critical value for identifying IO and AO, and C_{LS} is the critical value for the detection of LS. The results of the procedure of Chen and Liu (1993) are shown between brackets.

Case	C_{AI}	C_{LS}	PAR	RMSE(PAR)	PARF	RMSE(PARF)	$\hat{\sigma}_a$	RMSE($\hat{\sigma}_a$)
10	3.25	2.85	0.558	0.090	0.572	0.093	1.004	0.088
	(3.25)	(3.25)	(0.562)	(0.087)	(0.563)	(0.101)	(1.014)	(0.177)
11	3.25	2.85	0.560	0.090	0.575	0.094	0.985	0.058
	(3.25)	(3.25)	(0.562)	(0.087)	(0.563)	(0.101)	(1.014)	(0.177)
12	3.25	2.85	0.777	0.184	0.799	0.222	1.141	0.179
	(3.25)	(3.25)	(0.775)	(0.183)	(0.810)	(0.217)	(1.112)	(0.226)
	(3.00)	(3.00)	(0.775)	(0.183)	(0.796)	(0.212)	(1.046)	(0.216)
	(2.75)	(2.75)	(0.775)	(0.183)	(0.770)	(0.197)	(0.945)	(0.241)
13	3.25	2.85	0.415	0.213	0.565	0.103	0.986	0.078
	(3.25)	(3.25)	(0.416)	(0.212)	(0.548)	(0.117)	(0.960)	(0.181)
14	3.25	2.85	0.450	0.147	0.568	0.100	0.986	0.077
	(3.25)	(3.25)	(0.416)	(0.212)	(0.548)	(0.117)	(0.960)	(0.181)
15	3.25	2.85	0.690	0.113	0.740	0.154	1.049	0.095
	(3.25)	(3.25)	(0.692)	(0.111)	(0.738)	(0.152)	(1.103)	(0.213)
	(3.00)	(3.00)	(0.692)	(0.111)	(0.746)	(0.159)	(1.046)	(0.201)
	(2.75)	(2.75)	(0.692)	(0.111)	(0.757)	(0.171)	(0.939)	(0.222)

Tables 6 and 7 show the estimation results and the percentage of *correctly* detected outliers for cases 25 through 29 of table 2. These cases are not included in Chen and Liu (1993). The number of replications is 500. The critical value for identifying IO and AO is $C_{AI} = 3.25$, and the critical value for the LS is $C_{LS} = 2.85$. The range of the variable *number of outliers correctly detected* (table 7) goes from 0 to 4 outliers for all cases

Table 6 Performance of the procedure. Multiple outliers.

CASE	PAR	RMSE(PAR)	PARF	RMSE(PARF)	$\hat{\sigma}_a$	RMSE($\hat{\sigma}_a$)
25	0.043	0.567	0.571	0.096	1.045	0.160
26	0.111	0.504	0.567	0.099	1.064	0.184
27	0.186	0.431	0.578	0.093	1.040	0.138
28	0.037	0.575	0.563	0.102	1.045	0.178
29	0.179	0.436	0.572	0.095	1.029	0.115

Cases 25 through 29 are four adjacent additive outliers, two or more of which were correctly detected in a range of 66% to 84% of the cases. Only in 2 of the 2500 series the four outliers are correctly detected. In spite of this behavior, the sample mean of the estimated autoregressive parameter after removing the detected outliers is quite close to the true value (0.6); the most unfavourable case is 0.563. This estimation is better than the one obtained supposing no outliers (which in the most favorable case is 0.186).

Table 7. Number of correctly detected outliers with the proposed procedure for the cases from 25 through 29 (percentage).

CASE	0	1	2	3	4
25	1.8	32.2	42.8	23.2	0.0
26	1.4	26.6	52.4	19.4	0.2
27	0.4	18.8	72.4	8.4	0.0
28	1.2	29.2	43.4	26.0	0.2
29	0.4	15.2	74.4	10.0	0.0

In consequence, although the procedure does not correctly identify the type of the outliers, the sample mean of the estimation of ϕ at the end of the procedure is very close to the true value. These apparently contradictory results of incorrect detection but correct estimation of the parameters, can be explained bearing in mind the concept of *equivalent configurations*. That is, every sequence of outliers, being two or more adjacents, can lead to different *equivalent configurations* of outliers. Two configurations of outliers are equivalent when their effects on a time series are indistinguishable. For example, a configuration of an additive outlier of magnitude $-\omega$ at instant $t = T$ is equivalent to a configuration of two adjacent level shifts in $t = T$ and $t = T + 1$ of magnitudes ω and $-\omega$ respectively. In the same manner, two additive outliers in $t = T$ and $t = T + 1$ with magnitudes $-\omega_1$ and ω_2 respectively are equivalent to an innovational outlier and an additive outlier in the same instant when $1 < \theta < 0$. Finally, three adjacent additive outliers in $t = T, T + 1, T + 2$ are equivalent to an AO in $t = T$, an IO in $t = T + 1$ and an AO in $t = T + 2$, with different outlier magnitudes in each case.

For instance, in cases 25 to 29 we have an AR(1) with four additive outliers of similar size and alternating signs. Then we can write :

$$\omega I_t^{(T)} - \omega I_t^{(T+1)} + \omega I_t^{(T+2)} - \omega I_t^{(T+3)} \simeq \omega(1-\phi B)^{-1} I_t^{(T)} - \omega_1 I_t^{(T+1)} + \omega_2 I_t^{(T+2)} - \omega I_t^{(T+3)}$$

and if ϕ is not large so that $\phi^3 \simeq 0$ the expression on the left hand is an equivalent configurations to the original one with the sequence of outliers IO,AO,AO,AO and sizes $\omega_1 = \omega(\phi + 1)$ and $\omega_2 = \omega(1 - \phi^2)$. As here ϕ is 0.6 we expect that the left hand configuration can be found quite often. Thus, if the procedure finds three additive outliers we expect that the fourth can easily be IO and with a size given by $\omega(1 - \phi^2)$.

In the same way if two AO are correctly detected and the modulus of the different AO are the same, then we have equivalent configurations as :

$$\omega I_t^{(T)} - \omega I_t^{(T+1)} + \omega I_t^{(T+2)} - \omega I_t^{(T+3)} \simeq \omega(1-\phi B)^{-1} I_t^{(T)} - \omega_1 (1-\phi B)^{-1} I_t^{(T+1)}$$

that is IO,IO,AO,AO with $\omega_1 = \omega(\phi+1)$, $\omega_2 = \omega(1+\phi)$ and $\omega_3 = \omega(1-\phi^2(1+\phi))$. Other possible configurations are IO,AO,IO,AO with $\omega_1 = \omega(\phi + 1)$, $\omega_2 = \omega(1 - \phi^2)$ and $\omega_3 = \omega(1 + \phi(1 - \phi^2))$; AO,IO,IO,AO with $\omega_1 = \omega_2 = \omega(1 + \phi)$; IO,AO,-,AO with $\omega_1 = \omega(1 + \phi)$ and AO,IO,AO, with $\omega_1 = \omega(1 + \phi)$ respectively.

Cases 25 and 28 have four AO with the same modulus, but opposite sign. We show in table 8 for each case the percentage of runs for which a configurations equivalent to the original model is detected. Thus, we observe that a high proportion of the times the procedure has detected one of the equivalent configurations that behave approximately as four consecutive AO. The rest of the cases can be explained in a similar way.

Table 8. Percentage of equivalent configurations for three and two AO correctly detected (case 25 and case 28).

Eq. Configurations	Case 25	Case 28
IO,AO,AO,AO	14.66	14.62
AO,IO,AO,AO	48.28	53.08
Total	62.94	67.7
IO,IO,AO,AO	10.05	8.73
IO,AO,IO,AO	8.13	9.61
AO,IO,IO,AO	2.39	0.44
IO,AO,-,AO	29.18	34
AO,IO,AO,-	11.48	10.48
Total	61.23	63.26

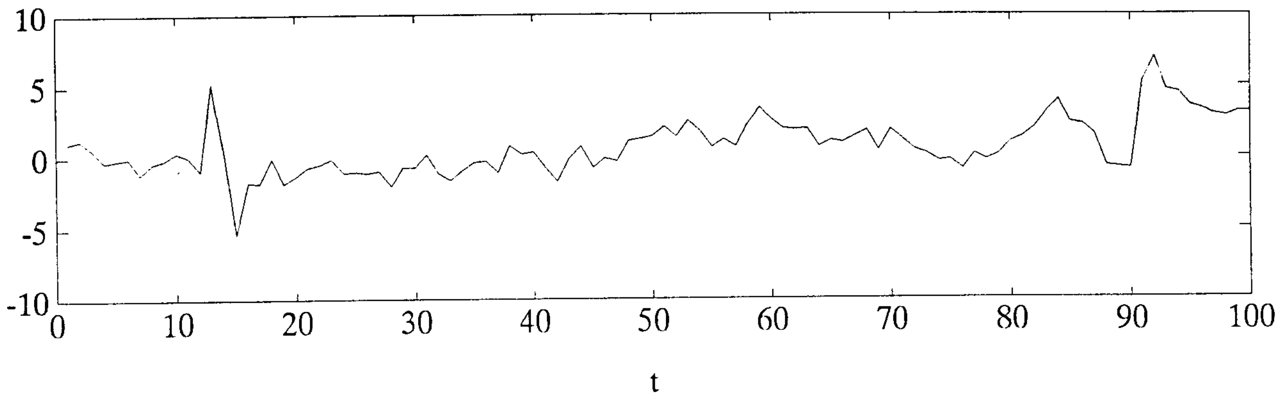
In summary, when dealing with a series that has adjacent outliers, we should take into account the correct detection as well as the multiple equivalent configurations that can be expected.

6.1 An Example

To illustrate how the proposed procedure works, we use the simulated series from the AR(1) model with parameter $\phi = 0.7$, and with $a_t \rightsquigarrow N(0, 1)$. It has been generated 150 observations with the

SCA Statistical System for computation, with the seed 250993. The first 50 observations have been eliminated. Thus, the sample size is $n = 100$. The estimated model is $(1 - 0.78B)z_t = a_t$ with $\hat{s}_R = 0.82$. This series is free of outliers and we have contaminated it with an AO at $t = 13$ of magnitude 5, an AO at $t = 15$ of magnitude -5, a LS at $t = 91$ of magnitude 4 and an AO at $t = 92$ of magnitude 3. Figure 2 shows the plot of the observed series.

Figure 2 Plot of the observed series.



The proposed procedure is described as follows:

In the stage 1, *Initial estimation of the model parameters*, we obtain that the maximum value of the measure of influence for a LS, $DL(t)$, occurs in t=91. An intervention model with a step variable in t=91 is built, resulting that its effect (4.1714) is significant. This effect is removed, and the measure for LS is built again, but now the estimated effect is no significant. Thus, the individual influence measure $D_{\hat{Z}}$ is calculated in every time point. The 10% of the observations, those with the highest values of influence measure, are corrected as additive outliers. Table 9 shows the time points and the values of $D_{\hat{Z}}$ for these observations.

Table 9. The individual measure of influence.

t	12	13	15	16	18
$D_{\hat{Z}}$	11.3	27.6	20.3	1.7	2.1
t	28	37	42	84	92
$D_{\hat{Z}}$	1.7	1.7	1.7	2.2	4.0

At the end of the stage 1 the estimation of the autoregressive parameter is $\hat{\phi} = 0.75$.

Stage 2, *Outliers detection*, starts with the observed series and the estimation of the parameters obtained in stage 1. The residuals and the $\hat{\lambda}_{I,t}$, $\hat{\lambda}_{A,t}$ and $\hat{\lambda}_{L,t}$ $(t = 1, \ldots, n)$ are now calculated. For every instant, $|\hat{\lambda}_{I,t}|$, $|\hat{\lambda}_{A,t}|$ are compared, and the maximum value of them is compared against $C_1 = 3$. On the other hand, the maximum value of $|\hat{\lambda}_{L,t}|$ is compared against $C_2 = 3$. From these comparisons, the procedure detects iteratively an AO in $t = 13$, a LS in $t = 91$, an AO in $t = 92$ and an AO at $t = 15$. The additive outliers and level shift 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 the stage 3, *Joint estimation*, we carry out the joint estimation (maximum likelihood) of the autoregressive parameters of the model and the effects of three AO in $t = 13, 15, 92$ and a LS at $t = 91$. The final model is:

$$(1 - \phi_1 B)(\tilde{z}_t - \omega_{13} I_t^{13} - \omega_{15} I_t^{15} - \omega_{91} S_t^{91} - \omega_{92} I_t^{92}) = a_t$$

The estimate parameters, their t values and the residual standard error are

Table 10. The estimate of parameters of the model.

Type of parameter	Value	t values	$\hat{s}_R$
Autoregressive	0.81	16.45	
Effect of the AO in t=13	5.51	8.79	
Effect of the AO in t=15	-4.66	-7.43	
Effect of the LS in t=91	5.15	7.16	
Effect of the AO in t=92	2.02	3.23	
			0.81

In order to look for more outliers, stage 2 starts again with the adjusted series from three AO in t=13, t=15 and t=92 (with estimated effects equal to 5.51, -4.66 and 2.02 respectively) and a LS in t=91 (with estimated effect equal to 5.15) and using the aforementioned autoregressive parameter. Since the procedure does not find more outliers, it finishes.

The estimated autoregressive parameters, the estimated effects of three additive outlier in t=13, 15 and 92 and a LS in t=91 are presented in table 10.

The procedure of Chen and Liu (1993) for the same series and with the choice C=3, correctly detects the three AO at t=13, 15 and 92, but at t=91 it detects an IO instead of the LS.

Appendix

A 95% Percentiles of $max \mid \hat{\lambda}_{i,t} \mid$ for IO and LS

Table I. Critical values to 95 % for $AR(1), AR(2)$ and $AR(3)$ models, where n is the sample size and nm is the number of model in table 1.

n	nm	IO	LS	nm	IO	LS	nm	IO	LS
30	1	2.89	2.34	8	2.89	2.89	15	2.92	3.08
30	2	2.89	2.46	9	2.91	2.70	16	2.93	2.44
30	3	3.00	2.70	10	2.94	2.31	17	2.96	2.64
30	4	3.01	2.50	11	2.97	2.86	18	2.92	2.70
30	5	2.90	2.15	12	2.95	2.33	19	2.89	2.62
30	6	2.88	2.16	13	2.92	2.26	20	2.88	2.25
30	7	2.86	2.12	14	3.02	2.68	21	2.85	2.06
50	1	3.10	2.51	8	3.14	2.99	15	3.12	3.13
50	2	3.07	2.52	9	3.13	2.92	16	3.09	2.65
50	3	3.21	2.77	10	3.11	2.49	17	3.06	2.58
50	4	3.15	2.95	11	3.12	2.90	18	3.10	2.83
50	5	3.07	2.30	12	3.22	2.51	19	3.10	2.65
50	6	3.12	2.30	13	3.11	2.29	20	3.14	2.40
50	7	3.09	2.24	14	3.09	2.72	21	3.11	2.29
100	1	3.29	2.57	8	3.42	2.99	15	3.34	3.18
100	2	3.42	2.98	9	3.34	2.90	16	3.30	2.72
100	3	3.36	2.66	10	3.38	2.64	17	3.38	2.78
100	4	3.39	3.06	11	3.37	3.13	18	3.39	2.90
100	5	3.38	2.62	12	3.45	2.63	19	3.35	2.81
100	6	3.36	2.47	13	3.38	2.65	20	3.36	2.62
100	7	3.30	2.60	14	3.43	2.85	21	3.45	2.33
150	1	3.48	2.77	8	3.50	2.99	15	3.42	3.25
150	2	3.52	2.70	9	3.42	2.97	16	3.57	2.88
150	3	3.54	2.82	10	3.52	2.55	17	3.51	2.85
150	4	3.55	3.04	11	3.62	3.04	18	3.54	2.92
150	5	3.57	2.52	12	3.55	2.64	19	3.57	2.74
150	6	3.58	2.61	13	3.44	2.49	20	3.57	2.56
150	7	3.42	2.53	14	3.57	2.85	21	3.46	2.35

The critical values C_i used in the detection of outliers have been obtained with multiple regression models from the critical values that are presented in this appendix. The response variables are, on the one hand, the critical values for IO and AO and, on the other hand, the critical values for LS. For the first model, the explicative variables are: the sample size, the type of outlier (additive or innovative) and the structure of the model. For the second model, the explicative variables are the sample size, the order of the model and the structure model.

B Relationship between $D_{\hat{Z}}(T)$ and $P_{\hat{Z}}(T)$

The vector of forecasts $\hat{\mathbf{Z}}_T^{(INT)}$ is obtained with the same parameters $\hat{\boldsymbol{\pi}}_{(T)}$ that are used in $\hat{\mathbf{Z}}_{(T)}$ (Peña 1987; and Ljung 1993 prove that the maximum likelihood function for the parameters $\boldsymbol{\pi}$ in an intervention model is approximately the same that the one obtained supposing an unobserved observation and interpolating with the inverse autocorrelation function.). Using $\pi(B)(z_t - \omega_A I_t^{(T)}) = a_t$, the vector of forecasts $\hat{\mathbf{Z}}_T^{(INT)}$ can be written as:

$$\hat{\mathbf{Z}}_T^{(INT)} = (\mathbf{X}_z - \hat{\omega}_A \mathbf{O})\hat{\boldsymbol{\pi}}_{(T)} + \hat{\omega}_A \mathbf{I}_t^{(T)},$$

where $\mathbf{O}$ is a matrix with the same dimension as $\mathbf{X}_z$ and has the form

$$\mathbf{O} = \begin{pmatrix} \mathbf{0} \\ -- \\ \mathbf{I}_h \\ -- \\ \mathbf{0} \end{pmatrix},$$

where $\mathbf{I}_h$ is the identity matrix of order h (the order of the model). $\mathbf{I}_h$ begins in the instant $T+1$. Thus, since $\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_T^{(INT)} = (\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_{(T)}) + \hat{\omega}_A(\mathbf{O}\hat{\pi}_{(T)} - \mathbf{I}_t^{(T)})$, the numerator of (7) can be written as:

$$\begin{aligned}(\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_T^{(INT)})'(\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_T^{(INT)}) &= (\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_{(T)})'(\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_{(T)}) \\ &+ \hat{\omega}_A^2(\mathbf{O}\hat{\pi}_{(T)} - I_t^{(T)})'(\mathbf{O}\hat{\pi}_{(T)} - \mathbf{I}_t^{(T)}) \\ &+ 2 \times \hat{\omega}_A(\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_{(T)})'(\mathbf{O}\hat{\pi}_{(T)} - \mathbf{I}_t^{(T)}). \qquad \text{(B-1)}\end{aligned}$$

Writing equation (B-1) in terms of the influence statistics, we have:

$$hD_{\hat{Z}}(T)\hat{\sigma}_a^2 = hP_{\hat{Z}}(T)\hat{\sigma}_a^2 + \hat{\omega}_A^2 \mathbf{B}_{(T)}'\mathbf{B}_{(T)} + 2 \times \hat{\omega}_A(\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_{(T)})'\mathbf{B}_{(T)}, \qquad \text{(B-2)}$$

where

$$\mathbf{B}_{(T)} = (\mathbf{O}\hat{\pi}_{(T)} - \mathbf{I}_t^{(T)}) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ -1 \\ \hat{\pi}_{1(T)} \\ \hat{\pi}_{2(T)} \\ \vdots \end{pmatrix},$$

and therefore,

$$\hat{\omega}_A^2 \mathbf{B}_{(T)}'\mathbf{B}_{(T)} = \hat{\omega}_A^2(1 + \hat{\pi}_{1(T)}^2 + \hat{\pi}_{2(T)}^2 + \ldots) = \hat{\omega}_A^2 \sum_{i=0}^{n-T} \hat{\pi}_{i(T)}^2,$$

with $\hat{\pi}_{0(T)}^2 = 1$. For the third term in (B-2), $\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_{(T)}$ is equivalent to

$$\mathbf{Z} - \mathbf{e} - (\mathbf{Z} - \mathbf{e}_{(T)}) = \mathbf{e}_{(T)} - \mathbf{e},$$

where $e_{(T)}$ are the residuals obtained filtering z_t for $\pi_{(T)}(B)$. Thus $e_{(T)} = \pi_{(T)}(B)z_t$, and the third term can be written as follows

$$\hat{\omega}_A(\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_{(T)})' B_{(T)} = \hat{\omega}_A(\mathbf{e}_{(T)} - \mathbf{e})' \begin{pmatrix} 0 \\ \vdots \\ 0 \\ -1 \\ \hat{\pi}_{1(T)} \\ \hat{\pi}_{2(T)} \\ \vdots \end{pmatrix},$$

and,

$$\hat{\omega}_A(\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_{(T)})' \mathbf{B}_{(T)} = \hat{\omega}_A \hat{\pi}_{(T)}(F)(e_T - e_{T(T)}).$$

Therefore, equation (B-2) can be written as:

$$hD_{\hat{Z}}(T)\hat{\sigma}_a^2 = hP_{\hat{Z}}(T)\hat{\sigma}_a^2 + \hat{\omega}_A^2 \sum_{i=0}^{n-T} \hat{\pi}_{i(T)}^2 + 2\hat{\omega}_A \hat{\pi}_{(T)}(F)(e_T - e_{T(T)}).$$

This equation can be related to the statistic for testing additive outliers ($\hat{\lambda}_{A,T} = \hat{\omega}_A \sqrt{\sum_{i=0}^{n-T} \hat{\pi}_{i,(T)}^2} / \hat{\sigma}_a$), obtaining

$$D_{\hat{Z}}(T) = P_{\hat{Z}}(T) - \frac{\hat{\lambda}_{A,T}^2}{h} + \frac{2\hat{\omega}_A}{h\hat{\sigma}_a^2} \hat{\pi}_T(F) e_T.$$

If $\hat{\omega}_0 = \hat{\pi}_{(T)}(F)e_T / \sum_{i=0}^{n-T} \hat{\pi}_{i(T)}^2$ is the estimator of ω obtained with the *wrong residuals* (those obtained supposing that there are not outliers), then

$$D_{\hat{Z}}(T) = P_{\hat{Z}}(T) - \frac{\hat{\lambda}_{A,T}^2}{h} + \frac{2\hat{\omega}_A \hat{\omega}_0}{h\hat{\sigma}_a^2} \sum_{i=0}^{n-T} \hat{\pi}_{i,(T)}^2,$$

and assuming that $(\hat{\omega}_A - \hat{\omega}_0)$ is small with respect to $\hat{\omega}_A^2$ this equation can be approximated by

$$D_{\hat{Z}}(T) \simeq P_{\hat{Z}}(T) + \frac{\hat{\lambda}_{A,T}^2}{h}.$$

C Relationship between $DL(T)$ and $PL(T)$

In this appendix we obtained the measure of the change in the parameters for a LS, $PL(T)$, and the relationship between the statistics $DL(T)$ and $PL(T)$. It can be shown that the ML estimates can be obtained from the expressions $\hat{\boldsymbol{\pi}} = (\mathbf{X}_z'\mathbf{X}_z)^{-1}\mathbf{X}_z'\mathbf{Z}$, and $\hat{\boldsymbol{\pi}}_L = (\hat{\mathbf{X}}_y'\hat{\mathbf{X}}_y)^{-1}\hat{\mathbf{X}}_y'\hat{\mathbf{Y}}$, where the matrix $\mathbf{X}_z$ has been described in section 4, $\hat{\mathbf{X}}_y$ is the matrix of estimated values for the real process that is unobserved in $t \geq T$, with $\hat{y}_t = z_t$ for $t < T$ and $\hat{y}_t = z_t - \hat{\omega}_L$ for $t \geq T$. The matrix $\mathbf{X}_z$ and the vector $\mathbf{Z}$ can be written in terms of $\hat{\mathbf{X}}_y$ and $\hat{\mathbf{Y}}$ respectively as:

$$\mathbf{X}_z = \hat{\mathbf{X}}_y + \hat{\omega}_L \mathbf{M}_L, \tag{C-1}$$

$$\mathbf{Z} = \hat{\mathbf{Y}} + \hat{\omega}_L \mathbf{V}_L, \tag{C-2}$$

where $\mathbf{M}_L$ is a matrix of zeros and ones that can be partitioned as $\mathbf{M}'_L = [\mathbf{M}'_L(1)\mathbf{M}'_L(2)]$, where $\mathbf{M}_L(1) = \mathbf{0}_{(T-h)\times h}$ and $\mathbf{M}_L(2)$ is a matrix of dimension $(n-T) \times h$, that has a lower triangular submatrix of ones (**TI**) and a submatrix of ones (**MU**). Thus, $\mathbf{M}'_L(2) = [\mathbf{TI}_{h\times h}; \mathbf{MU}_{h\times(n-h-T)}]$. The vector $\mathbf{V}_L$ can be written as $\mathbf{V}'_L = [\mathbf{V}'_L(1) \;\; \mathbf{V}'_L(2)]$, where $\mathbf{V}_L(1)$ is a column vector of dimension $(T-h) \times 1$ with zeros in every element except a one in the element $T-h$, and $\mathbf{V}_L(2)$ is a column vector of ones of dimension $(n-T) \times 1$. The matrices $\mathbf{X}_z$ and $\hat{\mathbf{X}}_y$ are partitioned in the same way as $\mathbf{M}_L$, and the vectors $\mathbf{Z}$ and $\hat{\mathbf{Y}}$ are partitioned like $\mathbf{V}_L$. The relationship between $\hat{\mathbf{X}}'_y\hat{\mathbf{X}}_y$ and $\mathbf{X}'_z\mathbf{X}_z$ can be obtained by using (C-1) as follows

$$\hat{\mathbf{X}}'_y\hat{\mathbf{X}}_y = \mathbf{X}'_z\mathbf{X}_z - \hat{\omega}_L(\mathbf{M}'_L(2)\mathbf{X}_z(2) + \mathbf{X}'_z(2)\mathbf{M}_L(2) - \hat{\omega}_L\mathbf{M}'_L(2)\mathbf{M}_L(2)) = \mathbf{X}'_z\mathbf{X}_z - \hat{\omega}_L\mathbf{L}_L,$$

therefore

$$\hat{\mathbf{X}}'_y\hat{\mathbf{X}}_y = \mathbf{X}'_z\mathbf{X}_z - \hat{\omega}_L\mathbf{L}_L \tag{C-3}$$

$$\mathbf{L}_L = \mathbf{M}'_L(2)\mathbf{X}_z(2) + \mathbf{X}'_z(2)\mathbf{M}_L(2) - \hat{\omega}_L\mathbf{M}'_L(2)\mathbf{M}_L(2). \tag{C-4}$$

The relationship between $\hat{\mathbf{X}}'_y\hat{\mathbf{Y}}$ and $\mathbf{X}'_z\mathbf{Z}$ is obtained using (C-1) and (C-2),

$$\hat{\mathbf{X}}'_y\hat{\mathbf{Y}} = (\mathbf{X}'_z - \hat{\omega}_L\mathbf{M}'_L)(\mathbf{Z} - \hat{\omega}_L\mathbf{V}_L) = \mathbf{X}'_z\mathbf{Z} - \hat{\omega}_L\mathbf{X}'_z\mathbf{V}_L - \hat{\omega}_L\mathbf{M}'_L\mathbf{Z} + \hat{\omega}_L^2\mathbf{M}'_L\mathbf{V}_L,$$

which can be written as

$$\hat{\mathbf{X}}'_y\hat{\mathbf{Y}} = \mathbf{X}'_z\mathbf{Z} - \hat{\omega}_L\mathbf{S}_L, \tag{C-5}$$

with

$$\mathbf{S}_L = \mathbf{X}'_z\mathbf{V}_L + \mathbf{M}'_L(2)\mathbf{Z}(2) - \hat{\omega}_L\mathbf{M}'_L(2)\mathbf{V}_L(2), \tag{C-6}$$

Substituting (C-3) and (C-5), in the expression of $\hat{\pi}_L = (\hat{\mathbf{X}}'_y\hat{\mathbf{X}}_y)^{-1}\hat{\mathbf{X}}'_y\hat{\mathbf{Y}}$, we get

$$(\mathbf{X}'_z\mathbf{X}_z - \hat{\omega}_L\mathbf{L}_L)\hat{\pi}_L = \mathbf{X}'_z\mathbf{Z} - \hat{\omega}_L\mathbf{S}_L,$$

and operating, it results

$$\hat{\pi}_L = \hat{\pi} - \hat{\omega}_L(\mathbf{X}'_z\mathbf{X}_z)^{-1}(\mathbf{S}_L - \mathbf{L}_L\hat{\pi}_L), \tag{C-7}$$

calling $\hat{\mathbf{E}}_L = \mathbf{S}_L - \mathbf{L}_L\hat{\pi}_L$, the relation between $\hat{\pi}$ and $\hat{\pi}_L$ becomes:

$$\hat{\pi} = \hat{\pi}_L + \hat{\omega}_L(\mathbf{X}'_z\mathbf{X}_z)^{-1}\hat{\mathbf{E}}_L$$

The influence measure for a LS, $PL(T)$, is defined as the Mahalanobis distance between the vectors $\hat{\pi}$ and $\hat{\pi}_L$. Thus we obtain

$$PL(T) = \frac{(\hat{\pi} - \hat{\pi}_L)'(\mathbf{X}'_z\mathbf{X}_z)(\hat{\pi} - \hat{\pi}_L)}{h\hat{\sigma}_a^2} = \frac{\hat{\lambda}_{L,T}^2}{1 + \sum \hat{l}_i^2} \frac{\hat{\mathbf{E}}'_L(\mathbf{X}'_z\mathbf{X}_z)^{-1}\hat{\mathbf{E}}_L}{h} \tag{C-8}$$

where $\hat{\lambda}_{L,T} = \hat{\omega}_L/\hat{\rho}_L\hat{\sigma}_a$ is the test statistic for a LS in $t = T$ with $\hat{\rho}_L = (1 + \hat{l}_1^2 + \hat{l}_2^2 + ... + \hat{l}_{n-T}^2)^{1/2}$, and $\hat{l}_i$ is i-th coefficient of $\hat{l}(B) = \hat{\pi}(B)/(1-B)$.

Next, we develop the relationship between the influence measures $DL(T)$ and $PL(T)$. Using $\pi(B)(z_t - \omega_L S_t^{(T)}) = a_t$, the vector of forecasts $\hat{\mathbf{Z}}_T^{(ILS)}$ can be written as:

$$\hat{\mathbf{Z}}_T^{(ILS)} = (\mathbf{X}_z - \hat{\omega}_L\mathbf{V})\hat{\pi}_{T,L} + \hat{\omega}_L\mathbf{S}_t^{(T)},$$

where $\mathbf{G}_T = \mathbf{V}\hat{\boldsymbol{\pi}}_{T,L}$ is the following vector

$$\mathbf{G}_T = \mathbf{V}\hat{\boldsymbol{\pi}}_{T,L} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \hat{\pi}_{1,L} \\ \hat{\pi}_{1,L} + \hat{\pi}_{2,L} \\ \vdots \\ \hat{\pi}_{1,L} + \hat{\pi}_{2,L} + \ldots + \hat{\pi}_{j,L} \\ \vdots \end{pmatrix}.$$

Thus

$$\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_T^{(ILS)} = (\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_{T,L}) + \hat{\omega}_L(\mathbf{G}_T - \mathbf{S}_t^{(T)}),$$

and

$$\begin{aligned}(\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_T^{(ILS)})'(\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_T^{(ILS)}) &= (\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_{T,L})'(\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_{T,L}) \\ &+ \hat{\omega}_L^2(\mathbf{G}_T - \mathbf{S}_t^{(T)})'(\mathbf{G}_T - \mathbf{S}_t^{(T)}) \\ &+ 2 \times \hat{\omega}_L(\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_{T,L})'(\mathbf{G}_T - \mathbf{S}_t^{(T)}).\end{aligned}$$

Writing this equation in terms of the influence statistics for LS we obtain:

$$hDL(T)\hat{\sigma}_a^2 = hPL(T)\hat{\sigma}_a^2 + \hat{\omega}_L^2\mathbf{L}_T'\mathbf{L}_T + 2 \times \hat{\omega}_L(\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_{T,L})'\mathbf{L}_T, \qquad \text{(C-9)}$$

where

$$\mathbf{L}_T = (\mathbf{G}_T - \mathbf{S}_t^{(T)}) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ -1 \\ -1 + \hat{\pi}_{1,L} \\ -1 + \hat{\pi}_{1,L} + \hat{\pi}_{2,L} \\ \vdots \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ -1 \\ \hat{l}_{1T} \\ \hat{l}_{2T} \\ \vdots \end{pmatrix},$$

and therefore,

$$\hat{\omega}_L^2\mathbf{L}_T'\mathbf{L}_T = \hat{\omega}_L^2(1 + \hat{l}_{1,T}^2 + \hat{l}_{2,T}^2 + \ldots) = \hat{\omega}_L^2 \sum_{i=0}^{n-T} \hat{l}_{i,T}^2,$$

with $\hat{l}_{0,T}^2 = 1$. In the third term of (C-9), $\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_{T,L}$ is equal to $\mathbf{e}_L - \mathbf{e}$, thus

$$\hat{\omega}_L(\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_{T,L})'\mathbf{L}_T = \hat{\omega}_L(\mathbf{e}_L - \mathbf{e})' \begin{pmatrix} 0 \\ \vdots \\ 0 \\ -1 \\ \hat{l}_{1,T} \\ \hat{l}_{2,T} \\ \vdots \end{pmatrix},$$

and operating, the third term of (C-9) results

$$2\hat{\omega}_L(\hat{\mathbf{Z}} - \hat{\mathbf{Z}}_{T,L})'\mathbf{L}_T = 2\hat{\omega}_L\hat{l}_T(F)(e_T - e_{T,L}),$$

and the equatión (C-9) can be written as:

$$hDL(T)\hat{\sigma}_a^2 = hPL(T)\hat{\sigma}_a^2 + \hat{\omega}_L^2 \sum_{i=0}^{n-T} \hat{l}_{i,T}^2 + 2\hat{\omega}_L\hat{l}_T(F)(e_T - e_{T,L}).$$

Using $\hat{\lambda}_{L,T} = \hat{\omega}_L\sqrt{\sum_{i=0}^{n-T} \hat{l}_{i,T}^2}/\hat{\sigma}_a$ (where $\hat{\omega}_L$ is the estimated effect of a LS) we have that:

$$DL(T) = PL(T) - \frac{\hat{\lambda}_{L,T}^2}{h} + \frac{2\hat{\omega}_L}{h\hat{\sigma}_a^2}\hat{l}_T(F)e_T,$$

and calling $\hat{\omega}_{0,L} = \hat{l}_T(F)e_T / \sum_{i=0}^{n-T} \hat{l}_{i,T}^2$ we have

$$DL(T) = PL(T) - \frac{\hat{\lambda}_{L,T}^2}{h} + \frac{2\hat{\omega}_L\hat{\omega}_{0,L}}{h\sigma_a^2} \sum_{i=0}^{n-T} \hat{l}_{i,T}^2.$$

and assuming that $(\hat{\omega}_L - \hat{\omega}_{0,L})$ is small compared to $\hat{\omega}_L^2$, this equation can be approximated by:

$$DL(T) \simeq PL(T) + \frac{\hat{\lambda}_{L,T}^2}{h},$$

and, as before, $DL(T)$ can be interpreted as the effect of the change of the parameters plus the effect of the particular intervention.

REFERENCES

Aoki, M. (1987). *State Space Modeling of Time.* Springer-Verlag.

Balke, N.S. (1993). "Detecting Level Shitfs in Time Series". *Journal of Bussiness and Economic Statistics,* 11, 1, 81–92.

Box, G.E.P. y Jenkins, G.M. (1976). *Time Series Analysis Forecasting and Control.* Holden Day.

Bruce, A.G. y 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. (1982). "General M-Estimates for Contaminated pth-order Autoregressive Processes: Consistency and Asymptotic Normality". *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 59, 491–504.

Bustos, O.H. y Yohai, V.J. (1986). "Robust Estimates for ARMA Models". *Journal of the American Statistical Association*, 81, 393, 155–168.

Chang, I. y 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. y Chen, C. (1988). "Estimation of Time Series Parameters in the Presence of Outliers". *Technometrics*, 30, 2, 193–204.

Chen, C. y Liu, L. (1993). "Joint Estimation of Model Parameters and Outlier Effects in Time Series". *Journal of the American Statistical Association*, 88, 421, 284–297.

Denby, L. y Martin, D. (1979). "Robust Estimation of the First-Order Autoregressive Parameter". *Journal of the American Statistical Association*, 74, 365, 140–146.

Fox, A.J. (1972). "Outliers in Time Series". *Journal of the Royal Statistical Society, B*, 34, 350–363.

Galbraith, J.W. y Zinde-Walsh, V. (1994). "A Simple Noniterative Estimator for Moving Average Models". *Biometrika*, 81, 1 , 143-155.

Harvey, A.C. (1981). *The Econometric Analysis of Time Series*. John Wiley.

Harvey, A.C. y Koopman, S.J. (1992)."Diagnostic checking of unobserved-components time series models". *Journal of Business and Economic Statistics*, 10, 377–389.

Kitagawa, G. (1987). "Non-Gaussian state-space modeling of nonstationary time series" (with discussion). *Journal of the American Statistical Association*, 82, 1032–1063.

Le, N.H., Martin, R.D., and 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, 436, 1504–1515.

Le, N.H., Raftery, A.E., and Martin, R.D. (1996). "Robust Bayesian Model Selection for Autoregressive Processes With Additive Outliers". *Journal of the American Statistical Association*, 91, 433, 123–131.

Lee, C.H. y Martin, R.D. (1982). "M-Estimates for ARMA Processes". Technical Report 23, Department of Statistics, University of Washington.

Ljung, G. (1993). "On Outlier Detection in Time Series". *Journal of the Royal Statistical Society, B*, 55, 2, 559–567.

Martin, R.D. (1980). "Robust Estimation for Autoregressive Models". *Directions in Time Series*, 228–254. Ed. D.R. Brillinger y G.C. Tiao. Institute of Mathematical Statistics.

Martin, R.D. (1981). "Robust Methods in Time Series". *Applied Time Series Analysis II*, 683–759. Ed. D.F. Findley. Academic Press.

Martin, R.D., Samarov, A. y Vandaele, W. (1983). "Robust Methods for ARIMA Models". *Applied Time Series Analysis of Economic Data*, 153–169. Ed. A. Zellner. Bureau of the Census.

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

Peña, D. (1987). "Measuring the Importance of Outliers in ARIMA Models". *New Perspectives in Theoretical and Applied Statistics,* 109–118. Ed. Puri et al. John Wiley.

Peña, D. (1990). "Influential Observations in Time Series". *Journal of Business and Economic Statistics,* 8, 2, –241.

Peña, D. (1991). "Measuring Influence in Dynamic Regression Models". *Technometrics,* 33, 1, 93–101.

Peña, D. and Guttman, I. (1989). "Optimal collapsing of mixture distributions in Robust Recursive Estimation" *Communications in Statistics, (Theory and Methods),* 18, 3, 817-834.

Taplin, R. S. (1993). "Robust Likelihood Calculation for Time Series". *Journal of the Royal Statistical Society, B,* 55, 4, 829-836.

Tsay, R.S. (1988). "Outliers, Level Shifts, and Variance Changes in Time Series". *Journal of Forecasting,* 7, 1–20.

West, M. (1981). "Robust sequential approximate Bayesian estimation". *Journal of the Royal Statistical Society, B,* 43, 157–166.

WORKING PAPERS 1997

Business Economics Series

97-18 (01) Margarita Samartín
"Optimal allocation of interest rate risk"

97-23 (02) Felipe Aparicio and Javier Estrada
"Empirical distributions of stock returns: european securities markets, 1990-95"

97-24 (03) Javier Estrada
"Random walks and the temporal dimension of risk"

97-29 (04) Margarita Samartín
"A model for financial intermediation and public intervention"

97-30 (05) Clara-Eugenia García
"Competing trough marketing adoption: a comparative study of insurance companies in Belgium and Spain"

97-31 (06) Juan-Pedro Gómez and Fernando Zapatero
"The role of institutional investors in international trading: an explanation of the home bias puzzle"

97-32 (07) Isabel Gutiérrez, Manuel Núñez Niekel and Luis R. Gómez-Mejía
"Executive transitions, firm performance, organizational survival and the nature of the principal-agent contract"

97-52 (08) Teresa García and Carlos Ocaña
"The role of banks in relaxing financial constraints: some evidence on the investment behavior of spanish firms"

97-59 (09) Rosa Rodríguez, Fernando Restoy and Ignacio Peña
"A general equilibrium approach to the stock returns and real activity relationship"

97-75 (10) Josep Tribo
"Long-term and short-term labor contracts versus long-term and short-term debt financial contracts"

Economics Series

97-04 (01) Iñigo Herguera and Stefan Lutz
"Trade policy and leapfrogging"

97-05 (02) Talitha Feenstra and Noemi Padrón
"Dynamic efficiency of environmental policy: the case of intertemporal model of emissions trading"

97-06 (03) José Luis Moraga and Noemi Padrón