Sociedad Española de Estadística e Investigación Operativa

Test

Volume 11, Number 2. December 2002

Robust Estimation in Time Series

Raúl P. Mentz and Carlos I. Martínez
Facultad de Ciencias Económicas,
Universidad Nacional de Tucumán and
CONICET (Argentina).

Sociedad de Estadística e Investigación Operativa
Test (2002) *Vol.* **11**, *No.* 2, pp. 385–404

Sociedad de Estadística e Investigación Operativa
Test (2002) *Vol.* **11**, *No.* 2, pp. 385–404

Robust Estimation in Time Series

Raúl P. Mentz* **and Carlos I. Martínez**
Facultad de Ciencias Económicas,
Universidad Nacional de Tucumán and
CONICET (Argentina).

Abstract

The main purpose of this work is to study empirically by means of simulations, the robustness of a set of proposals to estimate the parameters in the MA(1) time series model. The non-normal populations are mixtures of normal distributions, defined by $g(x) = pN(0,k) + (1-p)N(0,1)$, where the proportion of contamination most frequently used is $p = 0.10$ and k is the variance of the distribution used in the contamination; α is taken to be 0.90, which is close to the region of non-invertibility. Key results are that the estimation procedures used in the study provide good results in terms of biases in the estimation of the parameters, and that the biases are not changed when contaminated errors (mixtures) are considered. The estimation of the variance of the contaminated errors is also studied through simulations.

Key Words: Maximum likelihood estimation, contaminated errors, robustness, estimation of error variance, biases.

AMS subject classification: 62M10, 62F35

1 Introduction

Many estimation procedures used in time series analysis are deduced under restrictive assumptions. For example to estimate the parameters in ARMA models, frequently it is assumed that errors form a white noise process: they are independent (or at least non-correlated), with constant expected value (usually taken to be 0), and constant, finite variance. Under these assumptions, or similar ones, they are used computationally and their basic theoretical properties are studied; this applies, for example, to the method of moments or similar procedures (Burg's algorithm, for example), or some version of the least squares procedure.

*Correspondence to: Raúl P. Mentz, Casilla de Correo 209, (4000) Tucumán, Argentina, Fax 54 (381) 4364-105, phone 54 (381)4364-093. E-mail: rmentz@herrera.unt.edu.ar

Received: April 2001; Accepted: April 2002

With the additional assumption of normality, the method of maximum likelihood is frequently considered. Experiences in the use and analysis of this procedure are in general positive: asymptotic results are known covering various types of investigations, and studies conducted through simulations have frequently shown that the asymptotic results can be used with moderate and even small samples. In general, simulation studies use values generated by normal distributions.

The assumption of normality is often too demanding for applications. Several approaches have been used in the literature to deal with departures from this assumption in the case of ARMA models. We shall briefly review some of them.

One possibility is to apply methods to deal with outliers. The presence of outliers is an indication of the normality of the series. Fox (1972) discussed the idea of additive (AO, Type I) and innovations (IO, Type II) outliers. The former are the effect of external or exogenous causes and the latter of internal or endogenous causes. Chang et al. (1988) used the technique of intervention analysis (Box and Tiao (1975)) to deal with both kinds of outliers. Peña (1990) dealt with measuring the influence of outliers (Cook and Weisberg (1982)); in this approach observations are deleted and the effect of such deletions on the estimates is measured: a high influence means that the deletion strongly affects the estimates.

The treatment of outliers can also be approached through the use of robust statistical procedures. An exposition is Martin (1980) who dealt with autoregressive models. The basic idea is to replace the common weighting schemes (quadratic or absolute value), by more elaborate ones. A collection of these procedures is in Andrews et al. (1972), as will be discussed below.

Another approach is to use for the error term some non-normal distribution. Tikku et al. (2000) considered autoregressive models with errors terms modelled by a symmetric family of distributions, namely Student's t. These are known to have heavy tails for small degrees of freedom. These authors derive a modified maximum likelihood estimation procedure and show that it has good properties in estimation and testing problems.

An alternative related to the previous paragraph, is to use mixtures of normal distributions as models for the error term. In the frequently-quoted study Andrews et al. (1972), the behavior of a collection of as many as 68 point estimators of location was studied, but not in the context of time series. These authors used simulation and considered a wide variety of distributions. In particular, they used mixtures of normal distributions with equal expected values and differing variances, which are known in the literature as robustness models (Lindsay (1995), Section 1.3.12, Aitkin and Tunnicliffe (1980)).

In the present study we consider the first order moving average model with errors generated by robustness models. Through simulations we compare an iterative estimation procedure presented in Anderson and Mentz (1993b) with methods available in five well-known computer programs. Our main objective is to evaluate whether the use of mixture in the error term, affects the outcome of the procedures by introducing biases in the estimation of the moving average parameter.

2 The first-order moving average model and mixtures of two normal densities

The MA(1) (first-order moving average) model assumes that an observable time series y_t is generated by

$$y_t = u_t + \alpha u_{t-1}, \tag{2.1}$$

where the u_t are Gaussian white noise, that is, independent $N(0, \sigma^2)$ random variables. Instead of the parameters (α,σ^2), often the pair $(\sigma_0,\ \rho)$ is considered, where $\sigma_0 = \sigma^2(1+\alpha^2) = E(y_t^2)$ is the variance of y_t, and $\rho = \alpha/(1+\alpha^2) = E(y_t y_{t-1})/E(y_t^2)$ is the first-order autocorrelation coefficient. The relation between these parameters is $\rho = \alpha/(1+\alpha^2)$ and $\alpha = \left\{1-(1-4\rho^2)^{1/2}\right\}/(2\rho)$. The invertibility regions are $|\alpha| < 1$ and $|\rho| < 1/\{2\cos[\pi/(T+1)]\}$ (Anderson and Takemura (1986)). In general r is used to designate an estimator of ρ. The results are given in terms of ρ and only by exception in terms of α.

Instead of normal errors we consider the mixture of two normal densities, with the same expected values and different variances: $N(\theta, \sigma^2)$ y $N(\theta, k\sigma^2)$, where $k > 0$. Following Lindsay (1995), we write the mixture as $pN(\theta, k\sigma^2) + (1-p)N(\theta, \sigma^2)$, where $0 < p < 1$. Without loss of generality we take $\theta = 0$ y $\sigma = 1$.

The simulation procedure consists in generating pseudorandom independent numbers, uniformly distributed on $[0, 1]$, and to consider this interval divided by p. The experiment is interpreted as the selection of the

$N(0,k)$ density with relative expected proportion p, and density $N(0,1)$ with relative expected proportion $1-p$. When p is small, this is a model for the generation of a proportion p of outliers.

Hence, the model density can be written as

$$g(x;p,k) = p\frac{1}{\sqrt{2\pi k}}\exp\left\{-\frac{x^2}{2k}\right\} + (1-p)\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{x^2}{2}\right\}, \quad -\infty < x < \infty \tag{2.2}$$

where p and k are the model parameters.

The central moments of this density are obtained by integrating over x, so that they "may be expressed as weighted sums (using the same weights) of the expectations calculated using fixed numbers" (Andrews et al. (1972), Section 4D3). Hence, the density 2.1 has 0 expected value and variance

$$Var(X) = E(X^2) = pk + (1-p) \tag{2.3}$$

depending on the sampling fraction p and the scale factor k.

Given that the selection of the uniform random variables is done independently, it follows that the random variables with the mixture distribution are also independent. In the time series terminology, if X_t denotes the stochastic process generated in the indicated way, it constitutes a white noise process: independent random variables with 0 expected value and constant variance $pk + (1-p)$.

Following Andrews et al. (1972) (Section 5.4, Table 5-2), we take $k = 1, 9$ and 100: the first value gives random variables which are independent and identically distributed $N(0,1)$, and the other two correspond to standard deviations of 3 and 10, respectively.

In Figure 1 three densities are compared, namely: the standard normal $N(0,1)$, the mixture $0.5N(0,1) + 0.5N(0.9)$ (which has variance equal to 5, according to (2.3)), and the $N(0,5)$ density. The three densities are symmetric with respect to 0, but while the first and third ones are normal, the mixture is clearly non-normal, has "heavy tails", that is to say, it assigns high densities to values far from its 0 expected value.

Figures 2 and 3 compare de $N(0,1)$ density with mixtures having $k = 9$ and $k = 100$ respectively.

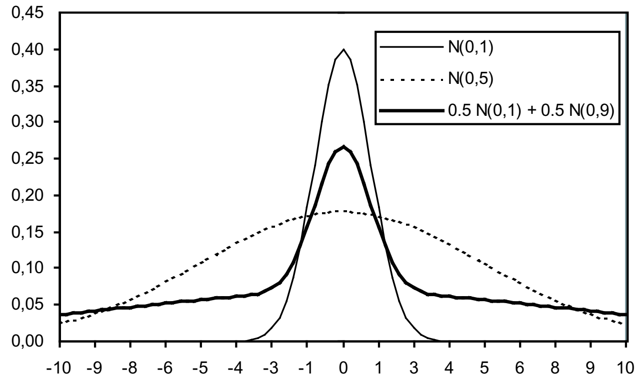

Figure 1: *Comparison of the densities* $N(0, 1)$ *and* $N(0, 5)$ *with the mixture* $0.5\ N(0, 1) + 0.5N(0, 9)$.

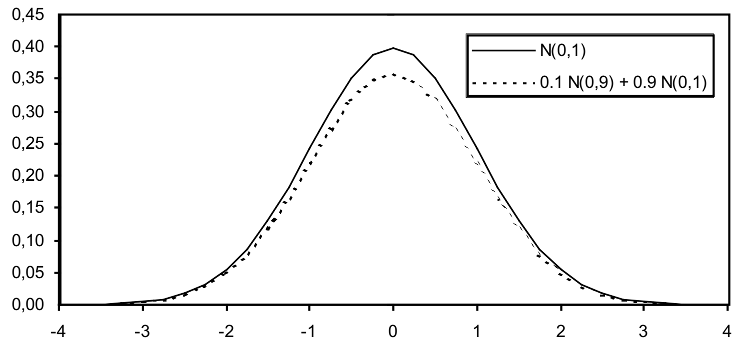

Figure 2: *Comparison of the mixtures used in Table 1.* $N(0, 1)$ *and mixture* $0.1\ N(0, 9) + 0.9\ N(0, 1)$

3 Estimation procedures in time series analysis

In terms of the parameters of the MA(1) model, the likelihood function of a vector of observations $y = (y_1, ..., y_T)'$ is

$$L(\sigma_0, \rho) = (2\pi\sigma_0)^{-T/2} |\boldsymbol{R}|^{-1/2} \exp\left\{-\boldsymbol{y}'\boldsymbol{R}^{-1}\boldsymbol{y}/(2\sigma_0)\right\} \tag{3.1}$$

where $\boldsymbol{R}$ is the $T \times T$ autocorrelation matrix: $\boldsymbol{R} = \boldsymbol{I} + \rho\boldsymbol{G}$, and $\boldsymbol{G}$ has 1's in its two diagonals adjacent to its main diagonal and 0's elsewhere.

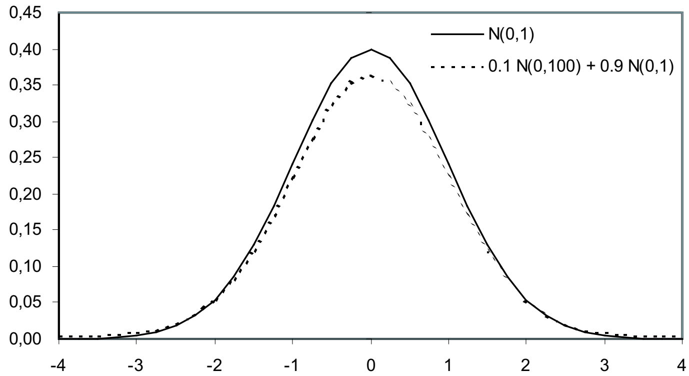

Figure 3: *Comparison of the mixtures used in Table 1.* $N(0,1)$ *and mixture* $0.1\ N(0,100)+0.9\ N(0,1)$

One iterative estimation procedures, identified as FORM1, is written

$$FORM1: \left\{t_{22}^{(i-1)}q_{10}^{(i-1)} - t_{21}^{(i-1)}q_{11}^{(i-1)}\right\} = r_i t_{20}^{(i-1)}q_{11}^{(i-1)} - t_{21}^{(i-1)}q_{10}^{(i-1)}, \quad (3.2)$$

where

$$q_{jk}^{(i-1)} = \boldsymbol{y}'\boldsymbol{R}_{i-1}^{-(j+1)}\boldsymbol{G}^k\boldsymbol{y}, \quad (3.3)$$

are quadratic forms, and

$$t_{jk}^{(i-1)} = tr\boldsymbol{R}_{i-1}^{-1}\boldsymbol{G}^k \quad (3.4)$$

are traces. The computational details and other analyses of this proposal can be found in Anderson and Mentz (1993a).

This iterative procedure is derived from the normal likelihood function when the method of scoring is used in the expansion of the log-likelihood. Other procedures are derived in Anderson and Mentz (1993b) by expanding the log-likelihood by the Newton-Raphson method, and also by using the two expansions with the concentrated likelihood functions. These procedures are mathematically equivalent, but they lead to different estimating equations, and hence often to different values of the estimates.

The estimation procedure defined by (3.2), (3.3), and (3.4) was compared in a simulation study with the following five:

1. *BMDPCls.* (Cls stands for conditional least squares) This is a preliminary estimation procedure, a variant of the method of moments, available

in BMDP (1990). The estimate of the main parameter minimizes the sum of squares appearing in the exponent of the Gaussian likelihood function, assuming certain initial values for the error terms. See Box and Jenkins (1970)(Chapter 7).

2. *BMDPBak*. (Bak stands for backcasting) In the previous procedure, the term corresponding to the Gaussian likelihood function is omitted. In the present one the full expression is considered. The initial values are not assumed to be fixed values, but they are "forecasted": this originates the expression backcasting (back-forecasting).

3. *S-PLUS*. Estimation procedure available in the package with this name. The procedure minimizes a likelihood function conditional on a set of initial values (Venables and Ripley (1997), Section 15.2). The program includes a set of alternative computations for the case of missing observations.

4. *ITSM.* (Interactive Time Series Modelling, Brockwell and Davis (1991)). This package includes two types of estimation procedures, preliminary and final, and this in turn can be least squares or Gaussian maximum likelihood; we only use the latter. The maximization is done by using the "innovations algorithm", which is a recursive procedure to compute the one-step ahead predictors and their mean square errors.

5. *MINITAB.* Estimation procedure for time series that follow an ARIMA model. MINITAB (1996).

The iterative procedure defined by (3.2)-(3.4) was studied by means of simulations in Anderson et al. (1996). In this paper the procedures mentioned above, were compared with the preliminary estimator of ρ given by the first-order sample autocorrelation r, and with the preliminary estimation procedure BMDPCls. Considering only pseudorandom normal numbers, the study by simulations detected that (3.4) operates quite satisfactorily for $T = 100$ when $\alpha = 0.30$, a value not in the non-invertibility region given by $|\alpha| > 1$; in this case, (3.4) substantially improves the simple estimator r, and BMDPCls provides results of a quality comparable with (3.4). However, when $\alpha = 0.90, T = 100$ was insufficient to justify the use of the asymptotic approximations known for the estimators and their standard errors. $T = 250$ was used and the fit to the asymptotic theory improved considerably. In the analysis the use of standard errors coming from the asymptotic theory was emphasized, as a means of evaluating the statistical significance of the simulation results; $n = 100$ replicates were done in each case.

The present paper has a structure similar to that in the cited publication, but here non-normal errors in model (2.1), generated by mixtures of normals as defined in (2.2) are used.

4 Description and analysis of the results

4.1 Simulation design

We consider samples sizes $T = 100$ or 250 and repetitions $n = 100$ or 500. We take the pair $(T, n) = (100, 500)$ as an example.

To generate the MA(1) observations with errors defined by (2.1), for each repetition the following steps produce the desired result:

1. Generate independent pseudorandom $N(0,1)$ numbers, by using Wolfram (1991). These are denoted $u_1, u_2, \ldots, u_{100}$.

2. Generate a value $\hat{p}$ from the uniform distribution on $(0,1)$.

3. Transform by defining $z_i^c = \sqrt{k} u_i, \quad \text{if } u_i \leq \hat{p}, \quad z_i^c = u_i$, if $u_i > \hat{p}$.

4. Model: $y_i = z_i^c + \alpha z_{i-1}^c, \quad i = 1, 2, ..., 100$.

5. Use the y_i to estimate the parameters of the MA(1) model.

6. Repeat $n = 500$ times.

To approximate the idea of independence, between each set of T observations, 50 were discarded. With a given set of y's, all the estimation procedures were applied; this design tries to control variability, at least in part: if the numbers were changed from the calculations for one method to the next one, differences due to these numbers will be added to the simulation process itself.

Calculations were done for three values of k, the variance of the contaminating distribution $N(0, k)$. We choose $k = 1, 9$ and 100, as was indicated in Section 2.

In the following subsection we describe the structure of Table 1, which contains many of the basic findings of our simulation study.

4.2 Structure of Table 1

Considering $r_i, i = 1, 2, ..., n$, the values coming from the n repetitions, Table 1 contains in its columns: (1)$m = m(r) = \sum_{i=1}^{n} r_i/n$, the estimates average, whose estimated standard error is $s/\sqrt{n}$, where s is in column (4); (2) $bi(r) = m - \rho$, the average bias, whose standard error is also $s/\sqrt{n}$, and is included in this column; (4)$s = s(r) = \left\{\sum_{i=1}^{n} (r_i - m)^2/(n-1)\right\}^{1/2}$, the standard deviation of the r_i, whose estimated asymptotic standard error is $s/\left\{2(n-1)\right\}^{1/2}$; (5) $as_m(r) = \left\{T^{-1}\left[1-\alpha^2\right]^3/\left[1+\alpha^2\right]^4\right\}^{1/2}$, the estimated asymptotic standard deviation of r; (6) $rs_\rho(r) = s/as_\rho(r)$, the empirical standard deviation of the r_i divided into its asymptotic standard deviation, whose estimated standard error is the standard error of r divided by the value of (5); (7) $mse(r) = s^2(r) + bi^2(r)$, the estimated mean squared error. The average estimate of α $m(\alpha)$, is also reported.

In columns (2) and (3), dividing the average estimates by their estimated standard errors, we obtain asymptotic tests of the null hypotheses that the true values are 0. For example, in the case of FORM1 and $k = 1, bi(r)/\left\{s/\sqrt{n}\right\} = -0.00093/0.00040 = -2.325$ indicates that for this sample of $n = 100$ replications, the bias is not significantly different from 0 at the 1% level. In column (6) we are interested in knowing if the sample quantities differ significantly from 1; for example, for BMDPBak and $k = 1, (1.01029 - 1)/0.00432 = 2.381$, which means that the hypothesis that the ratio does not differ from 1 is not rejected: the estimated standard deviation does not differ significantly from its asymptotic value.

4.3 Numerical results

Numerical results coming from simulation experiments are summarized in Table 1, which follows the structure described in Subsection 4.2.

The table has three parts, for $k = 1, 9$ and 100, respectively. In each part the following elements remain constant: (1) The MA(1) parameters, in particular $\alpha = 0.90$, to which $\rho = 0.49724$ corresponds; (2) Sample size $T = 100$; (3) The number of repetitions, $n = 100$, and (4) The contamination proportion $p = 0.10$ which is adequate for a robustness study, and is frequently used as such in the literature.

Results are given as functions of ρ and its estimators (denoted in general

Part a. $k = 1$

m(a)		(1) m(r)=m	(2) bi(r)	(3) rbi(r)	(4) s(r)	(5) asm(r)	(6) rsa(r)	(7) mse(r)
0,89739	FORM1	0,49631	-0,00093	-0,00187	0,00396	0,00264	1,56644	0,00303
			0,00040	0,00080	0,00028		0,00670	
0,91832	BMDP Bak	0,49749	0,00025	0,00050	0,00255	0,00183	1,01029	0,00280
			0,00026	0,00051	0,00018		0,00432	
0,88043	BMDP Cls	0,49451	-0,00273	-0,00548	0,00628	0,00338	2,48488	0,00356
			0,00063	0,00126	0,00045		0,01063	
0,90426	S-PLUS	0,49671	-0,00053	-0,00106	0,00292	0,00236	1,15699	0,00240
			0,00029	0,00059	0,00021		0,00495	
0,90410	ITSM	0,49671	-0,00053	-0,00107	0,00294	0,00236	1,16131	0,00240
			0,00029	0,00059	0,00021		0,00497	
0,91321	MINITAB	0,49727	0,00003	0,00006	0,00273	0,00201	1,07969	0,00276
			0,00027	0,00055	0,00019		0,00462	

Part b. $k = 9$

m(a)		(1) m(r)=m	(2) bi(r)	(3) rbi(r)	(4) s(r)	(5) asm(r)	(6) rsa(r)	(7) mse(r)
0,90829	FORM1	0,49705	-0,00019	-0,00038	0,00262	0,00220	1,03480	0,00243
			0,00026	0,00053	0,00019		0,00443	
0,92470	BMDP Bak	0,49780	0,00056	0,00113	0,00228	0,00160	0,90324	0,00285
			0,00023	0,00046	0,00016		0,00387	
0,88587	BMDP Cls	0,49492	-0,00232	-0,00467	0,00540	0,00313	2,13731	0,00308
			0,00054	0,00109	0,00038		0,00915	
0,91278	S-PLUS	0,49713	-0,00011	-0,00022	0,00266	0,00203	1,05136	0,00255
			0,00027	0,00053	0,00019		0,00450	
0,91228	ITSM	0,49712	-0,00012	-0,00024	0,00267	0,00205	1,05428	0,00254
			0,00027	0,00054	0,00019		0,00451	
0,92043	MINITAB	0,49758	0,00034	0,00069	0,00256	0,00175	1,01198	0,00290
			0,00026	0,00051	0,00018		0,00433	

Part c. $k = 100$

m(a)		(1) m(r)=m	(2) bi(r)	(3) rbi(r)	(4) s(r)	(5) asm(r)	(6) rsa(r)	(7) mse(r)
0,91111	FORM1	0,49708	-0,00016	-0,00032	0,00289	0,00209	1,14333	0,00273
			0,00029	0,00058	0,00021		0,00489	
0,93017	BMDP Bak	0,49790	0,00066	0,00133	0,00238	0,00142	0,94122	0,00304
			0,00024	0,00048	0,00017		0,00403	
0,89318	BMDP Cls	0,49394	-0,00330	-0,00665	0,01617	0,00281	6,39648	0,01287
			0,00162	0,00325	0,00115		0,02737	
0,91593	S-PLUS	0,49724	0,00000	0,00001	0,00284	0,00191	1,12333	0,00284
			0,00028	0,00057	0,00020		0,00481	
0,91562	ITSM	0,49723	-0,00001	-0,00001	0,00285	0,00192	1,12844	0,00285
			0,00029	0,00057	0,00020		0,00483	
0,92601	MINITAB	0,49772	0,00048	0,00097	0,00268	0,00156	1,05984	0,00316
			0,00027	0,00054	0,00019		0,00454	

Table 1: *Simulation results with the MA(1) model with parameter* $\alpha = 0.90$, ($\rho = 0.49724$), $T = 100$, *contamination of* $p = 0.10$, *and 3 values of* k

by r). Estimation results in terms of α are also presented to facilitate understanding.

Each part has 6 rows corresponding to the estimation procedures. The contents of each part was described in Subsection 4.2.

The main observations coming from the analysis of this information, are the following:

1.*Bias in the Estimation of* ρ. Column (1) contains the estimates average of ρ, column (2) the bias in the estimation of this parameter and (3) the relative bias. In column (1) we observe that the 5 procedures, except for BMDPCls, show non-significant biases, since these biases, divided by the corresponding standard errors (printed in parenthesis) are (approximately) less than 2. BMDPCls shows a significant bias, which is in agreement with its condition of being a preliminary estimation procedure.

These remarks are valid even when errors contaminated with $N(0, 9)$ and $N(0, 100)$ are used.

It is interesting to note the proximity of the results obtained by use of the programs S-PLUS and ITSM. This is understandable, since maximum likelihood estimation procedures to be used with ARMA models are widely available in the current literature.

In conclusion, the 5 programs provide good results in terms of biases in the estimation of ρ, and these biases do not change when we consider contaminated errors (mixtures) in the MA(1) model.

2. *Estimating the Variance of ρ's Estimators.* Column (4) contains the estimates of $\mathrm{Var}(r)$ for the given methods. These values can be compared with those coming from the asymptotic theory, as given in column (5). There exists a clear similarity between these two values, except in the case of BMDPCls (which was already discussed) and for FORM1 when $k = 1$.

The standard error of $s(r)$ in column (4) is computed, as indicated in Subsection 4.1, by means of the asymptotic expression $s/\sqrt{2(n-1)}$. For $k = 100$ and FORM1, this ratio is 0.00021, which is exactly the value in the table. In general, except for BMDPCls, the asymptotic approximations work very satisfactorily, with or without contamination.

To complete this section, we now compare the results obtained in Anderson et al. (1996) with those in the present paper. Table 2 summarizes these results.

We observe that the estimates coming from FORM1 are, in general, smaller than those of the other methods. This is due, in part at least, to the fact that this procedure, being an exact maximum likelihood, forces the estimates to be less than 1 in absolute value. This, in turn, comes from the fact that the likelihood function takes the same values in a given value of α and in its reciprocal. See, for example, Anderson and Mentz (1980).

Results in Table 2 show that the pairs of columns are quite similar,

MA(1), $\alpha = 0,90$ ($\rho = 0.49724$), $T = 100, n = 100, k = 1$				
Figures of Merit	FORM 1		BMDP Bak	
	This Paper	Anderson	This Paper	Anderson
M(r)=m	0.49631	0.49682	0.49749	0.49750
Bi(r)	-0.00093	-0.00042	0.00025	0.00026
	(0.00040)	(0.0041)	(0.00026)	
Rbi(r)	-0.00187	-0.00084	0.00050	0.00052
	(0.00080)	(0.00082)	(0.00051)	
S(r)	0.00396	0.00406	0.00255	0.00367
	(0.00028)	(0.00029)	(0.00018)	
Rsa(r)	0.00264	0.00282	0.00183	0.00234
Mse(r)	1.56644	1.605	1.01029	1.452
	(0.00670)	(0.115)	(0.0432)	

Table 2: *Comparison between simulation results in the present paper and those in Anderson et al. (1996)*

except for the estimated mean square error for procedure BMDPBak. A consequence of this similarity, is that in both works, biases and relative biases in the estimation of ρ are not significantly different from 0. The differences among estimated means square errors are due to the differences among the estimates of variances, since the contributions of the squared biases are small.

4.4 Other numerical results

Table 5 is similar to Table 1, except for the following: (1) Only 4 programs were analyzed, since BMDPCls was excluded (its behavior was inferior to that of BMDPBak), and ITSM (it was similar to S-PLUS, which is easier to use computationally); (2) $\alpha = 0.90$ ($\rho = 0.49724$) were used again, sample size $T = 100$ and three values of k $(1, 9$ and $100)$; (3) $n = 500$ repetitions were done and the contamination percentage was $p = 0.50$: this value is not associated with the usual ideas in robustness studies, but it was used as an extreme value to analyze its effects on the results.

A comparison of the results in Tables 1 and 5 is made in Table 7. The main observations stemming from this table are the following.

1. Estimates coming from FORM1 are in general smaller than those of the other methods. See Subsection 4.3.

$k = 1$

N^o	Rep	T	α	p	BMDP (CLS)	BMDP (BAK)	ITSM
1	100	100	0.90	0.10	1.0442	1.0081	1.0014
					(0.1711)	(0.1596)	(0.1573)
2	500	100	0.90	0.10	1.0367	0.9964	
					(0.1508)	(0.1384)	
3	500	250	0.90	0.10	1.0139	0.9952	
					(0.0926)	(0.0895)	
4	100	100	0.40	0.10	0.9895	09864	
					(0.1337)	(0.1336)	
				0.20	0.9946	0.9914	
					(0.1447)	(0.1436)	
				0.30	1.0085	1.0055	
					(0.1421)	(0.1415)	
				0.40	1.0051	0.9999	
					(0.1441)	(0.1422)	
				0.50	0.9875	0.9841	
					(0.1466)	(0.1461)	
				0.75	0.9634	0.9609	
					(0.1287)	(0.1286)	
5	100	100	0.40	0.05	1.0200	1.0163	
					(0.1486)	(0.1484)	
				0.10	1.0203	1.0161	
					(0.1484)	(0.1483)	
				0.20	0.9774	0.9743	
					(0.1342)	(0.1335)	
				0.30	1.0055	1.0015	
					(0.1308)	(0.1307)	
				0.40	1.0239	1.0180	
					(0.1482)	(0.1460)	
				0.50	0.9683	0.9648	
					(0.1467)	(0.1463)	
				0.75	1.0157	1.0107	
					(0.1582)	(0.1570)	

$k = 9$

N^o	Rep	T	α	p	BMDP (CLS)	BMDP (BAK)	ITSM
1	100	100	0.90	0.10	1.9086	2.0092	1.8302
					(0.5693)	(1.7823)	(0.5488)
2	500	100	0.90	0.10	1.8544	1.7844	
					(0.5038)	(0.4838)	
3	500	250	0.90	0.10	1.8267	1.7964	
					(0.3037)	(0.2970)	
4	100	100	0.40	0.10	1.8295	1.8249	
					(0.4508)	(0.4500)	
				0.20	2.5811	2.5711	
					(0.6883)	(0.6898)	
				0.30	3.4589	3.4450	
					(0.7483)	(0.7468)	
				0.40	4.9258	4.9052	
					(1.0339)	(1.0298)	
				0.50	5.0561	5.0444	
					(0.9219)	(0.9170)	
				0.75	6.9365	6.8488	
					(1.1903)	(1.1017)	

Continued

Table 3: *Simulation results with the MA(1) model, estimation of the variance of the error term defined as a mixture: average and standard deviation.*

Table 3. Continued

N^o	Rep	T	α	p	BMDP (CLS)	BMDP (BAK)	ITSM
5	100	100	0.40	0.05	1.4578	1.4535	
					(0.4370)	(0.4377)	
				0.10	1.8581	1.8523	
					(0.5011)	(0.5009)	
				0.20	2.5462	2.5383	
					(0.6137)	(0.6130)	
				0.30	3.4686	3.4549	
					(0.7846)	(0.7794)	
				0.40	4.3917	4.3681	
					(0.9300)	(0.9200)	
				0.50	4.8003	4.7832	
					(1.0838)	(1.0831)	
				0.75	6.9926	6.9545	
					(1.2268)	(1.2180)	

$k = 100$

N^o	Rep	T	α	p	BMDP (CLS)	BMDP (BAK)	ITSM
1	100	100	0.90	0.10	11.6665	11.2929	11.2742
					(6.2579)	(6.1470)	(6.0732)
2	500	100	0.90	0.10	11.1083	10.7577	
					(5.4224)	(5.3152)	
3	500	250	0.90	0.10	11.0699	10.9078	
					(3.2999)	(3.2530)	
4	100	100	0.40	0.05	6.4369	6.4230	
					(4.6777)	(4.6791)	
				0.10	11.4051	11.3766	
					(5.4712)	(5.4687)	
				0.20	20.4162	20.3548	
					(6.8408)	(6.8412)	
				0.30	31.4441	31.3293	
					(8.9559)	(8.8902)	
				0.40	42.6938	42.4690	
					(10.4204)	(10.3216)	
				0.50	48.4238	48.2631	
					(12.2090)	(12.2041)	
				0.75	74.9865	74.5799	
					(13.6208)	(13.5185)	

$k = 1$

N^o	Rep	T	α	p	BMDP (CLS)	BMDP (BAK)	ITSM
1	100	100	0.90	0.10	1.0442	1.0081	1.0014
2	500	100	0.90	0.10	1.0367	0.9964	
3	500	250	0.90	0.10	1.0139	0.9952	
4	100	100	0.40	0.10	0.9895	09864	
				0.20	0.9946	0.9914	
				0.30	1.0085	1.0055	
				0.40	1.0051	0.9999	
				0.50	0.9875	0.9841	
				0.75	0.9634	0.9609	
5	100	100	0.40	0.05	1.0200	1.0163	
				0.10	1.0203	1.0161	
				0.20	0.9774	0.9743	
				0.30	1.0055	1.0015	
				0.40	1.0239	1.0180	
				0.50	0.9683	0.9648	
				0.75	1.0157	1.0107	

$k = 9$

N^o	Rep	T	α	p	BMDP (CLS)	BMDP (BAK)	ITSM
1	100	100	0.90	0.10	1.9086	2.0092	1.8302
2	500	100	0.90	0.10	1.8544	1.7844	
3	500	250	0.90	0.10	1.8267	1.7964	
4	100	100	0.40	0.10	1.8295	1.8295	
				0.20	2.5811	2.5811	
				0.30	3.4589	3.4450	
				0.40	4.9258	4.9052	
				0.50	5.0561	5.0444	
				0.75	6.9365	6.8488	
5	100	100	0.40	0.05	1.4578	1.4535	
				0.10	1.8581	1.8523	
				0.20	2.5462	2.5383	
				0.30	3.4686	3.4549	
				0.40	4.3917	4.3681	
				0.50	4.8003	4.7832	
				0.75	6.9926	6.9545	

$k = 100$

N^o	Rep	T	α	p	BMDP (CLS)	BMDP (BAK)	ITSM
1	100	100	0.90	0.10	11.6665	11.2929	11.2742
2	500	100	0.90	0.10	11.1083	10.7577	
3	500	250	0.90	0.10	11.0699	10.9078	
44	100	100	0.40	0.05	6.4369	6.4230	
				0.10	11.4051	11.3766	
				0.20	20.4162	20.3548	
				0.30	31.4441	31.3293	
				0.40	42.6938	42.4690	
				0.50	48.4238	48.2631	
				0.75	74.9865	74.5799	

Table 4: Simulation results with the MA(1) model, estimation of the variance of the error term defined as a mixture: summary of average estimates.

Part a. $k = 1$

m(a)		(1) m(r)=m	(2) bi(r)	(3) rbi(r)	(4) s(r)	(5) asm(r)	(6) rsa(r)	(7) mse(r)
0,90230	FORM1	0,49641	-0,00083	-0,00166	0,00457	0,00109	4,03951	0,00374
			0,00020	0,00041	0,00014		0,00770	
0,91847	BMDP Bak	0,49722	-0,00002	-0,00003	0,00411	0,00081	3,63660	0,00409
			0,00018	0,00037	0,00013		0,00693	
0,90767	S-PLUS	0,49648	-0,00076	-0,00153	0,00472	0,00099	4,17675	0,00396
			0,00021	0,00042	0,00015		0,00796	
0,91657	MINITAB	0,49709	-0,00015	-0,00029	0,00432	0,00084	3,81777	0,00417
			0,00019	0,00039	0,00014		0,00325	

Part b. $k = 9$

m(a)		(1) m(r)=m	(2) bi(r)	(3) rbi(r)	(4) s(r)	(5) asm(r)	(6) rsa(r)	(7) mse(r)
0,90505	FORM1	0,49661	-0,00063	-0,00126	0,00426	0,00104	3,76911	0,00364
			0,00019	0,00038	0,00013		0,00718	
0,92158	BMDP Bak	0,49740	0,00016	0,00033	0,00387	0,00076	3,42288	0,00403
			0,00017	0,00035	0,00012		0,00652	
0,90920	S-PLUS	0,49670	-0,00054	-0,00108	0,00426	0,00097	3,77224	0,00373
			0,00019	0,00038	0,00013		0,00719	
0,91869	MINITAB	0,49725	0,00001	0,00002	0,00412	0,00081	3,64629	0,00413
			0,00018	0,00037	0,00013		0,00311	

Part c. $k = 100$

m(a)		(1) m(r)=m	(2) bi(r)	(3) rbi(r)	(4) s(r)	(5) asm(r)	(6) rsa(r)	(7) mse(r)
0,90501	FORM1	0,49669	-0,00055	-0,00111	0,00378	0,00104	3,34737	0,00323
			0,00017	0,00034	0,00012		0,00638	
0,92423	BMDP Bak	0,49753	0,00029	0,00057	0,00335	0,00072	2,96039	0,00363
			0,00015	0,00030	0,00011		0,00564	
0,91084	S-PLUS	0,49686	-0,00038	-0,00076	0,00370	0,00094	3,27483	0,00332
			0,00017	0,00033	0,00012		0,00624	
0,92080	MINITAB	0,49739	0,00015	0,00030	0,00354	0,00078	3,13142	0,00369
			0,00016	0,00032	0,00011		0,00267	

Table 5: Simulation results with the MA(1) model with parameter $\alpha = 0.90$ ($\rho =$ 0.49724) $T = 100, n = 500$*, contamination of* $p = 0.50$ *and 3values of* k

2. For $p = 0.10$, relative biases are non-significant at the 5% level, in all cases of Table 7 (which includes the two columns headed this paper in Table 1), except FORM1 when $k = 1$.

3. Estimates of $S(r)$ range from 0.00228 to 0.00292 for all procedures and values of k, except that FORM1 has the value 0.00396 for $k = 1$.

4. As a consequence of these observations, the estimated mean square error rme(r) takes values similar for the various procedures, except FORM1 as indicated.

5 Final comments,summary and conclusions

The main objective of this study is to perform an empirical analysis by means of simulations, the robustness of different proposals to estimate the parameters in the MA(1) time series model. The non-normal populations that we consider are mixtures of normal densities, defined in general by $g(x) = pN(0,k) + (1-p)N(0,1)$, where the contamination proportion is taken to be $p = 0, 10$ or 0.50, the variance of the contaminating normal distribution $k = 1, 9$ or 100 (Andrews et al. (1972)), and the estimation procedures are Cls (conditional least squares) and Bak ("backcasting") of BMDP (BMDP (1990)), SPLUS (Venables and Ripley (1997)), ITSM (Brockwell and Davis (1991)) and MINITAB (MINITAB (1996)).

The other components of the simulations program are: the MA(1) model parameter is taken to be $\alpha = 0.90$, except in Tables 3 and 4 where $\alpha = 0.40 : 0.90$ is a value close to the region of non-invertibility ($|\alpha| < 1$ for the MA(1) to be invertible into an infinite autoregression), sample size is $T = 100$, while some partial experiments were done with $T = 250$, and the number of repetitions is $n = 100$ or 500.

The results obtained in the estimation of α or ρ are presented with the format designed in Anderson et al. (1996), which is a simulation study using only FORM1 and BMDPBak.

The empirical results (Table 1) show that the programs (except for BMDPCls, as expected) give satisfactory results in terms of biases in the estimation of ρ, and that these biases do not change when contaminated errors (mixtures) are used in the MA(1) model. These results are compared, partially at least, with those in Anderson et al. (1996), having found a great deal of similarity among them, also as expected (Table 2); this similarity exists even when the sources of generation of the pseudorandom numbers are changed.

Since the error variance changes when we consider different mixtures, and is given by $\mathrm{Var}(u_t) = pk + (1-p)$, Tables 3 and 4 show the results of some experiments done with BMDPBak. This procedure is selected because it provides complete results, and it improves the estimation by BMDPCls. The following values are taken: $\alpha = 0.40, T = 100$ or $250, n =$ 100 or 500 and $p = 0.05, 0.10, 0.20, 0.30, 0.40, 0.50$ and 0.75. A summary of these results, compared with the corresponding theoretical or expected values, constitutes Table 6.

The good agreement of the empirical results with the underlying theory is observed.

Finally, in the experiment done by changing p from 0.10 to 0.50 (and n from 100 to 500), good stability of the results is observed.

Contamination (p)	$k = 1$ Var(u_t)	$k = 1$ Estimation BMDPBak	$k = 9$ Var(u_t)	$k = 9$ Estimation BMDPBak	$k = 100$ Var(u_t)	$k = 100$ Estimation BMDPBak
0.05	1.00	1.0163	1.40	1.4535	5.90	6.4230
0.10	1.00	1.0161	1.80	1.8523	10.90	11.3766
0.20	1.00	0.9743	2.60	2.5383	20.80	20.3548
0.30	1.00	1.0015	3.40	3.4549	30.70	31.3293
0.40	1.00	1.0180	4.20	4.3681	40.60	42.4690
0.50	1.00	0.9648	5.00	4.7832	50.50	48.2631
0.75	1.00	1.0107	7.00	6.9545	75.25	74.5799

Table 6: *Comparison between simulation with the MA(1) model, in estimating the variance of the error term defined by a mixture: comparison between empirical averages and theoretical Values*

References

AITKIN, M. and TUNNICLIFFE, W. G. (1980). Mixtures models, outliers and the EM algorithm. *Technometrics*, 22:325–331.

ANDERSON, T. W. and MENTZ, R. P. (1980). On the structure of the likelihood function of autoregressive and moving average models. *Journal of Time Series Analysis*, 1:83–94.

ANDERSON, T. W. and MENTZ, R. P. (1993a). Evaluation of quadratic forms and traces for iterative estimation in first-order moving average models. *Communications in Statistics: Theory and Methods*, 22:931–963.

ANDERSON, T. W. and MENTZ, R. P. (1993b). *Iterative Procedures for Exact Maximun Likelihood Estimation in the First-order Gaussian Moving Average Model*. A Raghu Raj Bahadur Fetchrift. Wiley Eastern Limited.

ANDERSON, T. W., MENTZ, R. P., JARMA, N. M., and MARTÍNEZ, C. I. (1996). Simulations of iterative procedures for maximum likelihood estimation in MA(1) models. *Communications in Statistics: Simulation and Computation*, 25:851–865.

ANDERSON, T. W. and TAKEMURA, A. (1986). Why do noninverrible estimated moving averages occur? *Journal of Time Series Analysis*, 7:235–254.

Part a. $k = 1$

Item	Form1		BMDBPBak		S-PLUS		MINITAB	
	$p = 0.10$	$p = 0.50$	$p = 0.10$	$p = 0.50$	$p = 0.10$	$p = 0.50$	$p = 0.10$	$p = 0.50$
	$n = 100$	$n = 500$	$n = 100$	$n = 500$	$n = 100$	$n = 500$	$n = 100$	$n = 500$
m(r) = m	0,49631	0,49641	0,49749	0,49722	0,49671	0,49648	0,49727	0,49709
bi(r)	-0,00093	-0,00083	0,00025	-0,00002	-0,00053	-0,00076	0,00003	-0,00015
	-0,0004	-0,0002	-0,00026	-0,00018	-0,00029	-0,00021	-0,00027	-0,00019
rbi(r)	-0,00187	-0,00166	0,00050	-0,00003	-0,00106	-0,00153	0,00006	-0,00029
	-0,0008	-0,00041	-0,00051	-0,00037	-0,00059	-0,00042	-0,00055	-0,00039
s(r)	0,00396	0,00457	0,00255	0,00411	0,00292	0,00472	0,00273	0,00432
	-0,00028	-0,00014	-0,00018	-0,00013	-0,00021	-0,00015	-0,00019	-0,00014
asm(r)	0,00264	0,00109	0,00183	0,00081	0,00236	0,00099	0,00201	0,00084
rsa(r)	1,56644	4,03951	1,01029	3,63660	1,15699	4,17675	1,07969	3,81777
	-0,0067	-0,0077	-0,00432	-0,00693	-0,00495	-0,00796	-0,00462	-0,00325
mse(r)	0,00303	0,00374	0,00280	0,00409	0,00240	0,00396	0,00276	0,00417

Part b. $k = 9$

Item	Form1		BMDBPBak		S-PLUS		MINITAB	
	$p = 0.10$	$p = 0.50$	$p = 0.10$	$p = 0.50$	$p = 0.10$	$p = 0.50$	$p = 0.10$	$p = 0.50$
	$n = 100$	$n = 500$	$n = 100$	$n = 500$	$n = 100$	$n = 500$	$n = 100$	$n = 500$
m(r) = m	0,49705	0,49661	0,49780	0,49740	0,49713	0,49670	0,49758	0,49725
bi(r)	-0,00019	-0,00063	0,00056	0,00016	-0,00011	-0,00054	0,00034	0,00001
	-0,00026	-0,00019	-0,00023	-0,00017	-0,00027	-0,00019	-0,00026	-0,00018
rbi(r)	-0,00038	-0,00126	0,00113	0,00033	-0,00022	-0,00108	0,00069	0,00002
	-0,00053	-0,00038	-0,00046	-0,00035	-0,00053	-0,00038	-0,00051	-0,00037
s(r)	0,00262	0,00426	0,00228	0,00387	0,00266	0,00426	0,00256	0,00412
	-0,00019	-0,00013	-0,00016	-0,00012	-0,00019	-0,00013	-0,00018	-0,00013
asm(r)	0,00220	0,00104	0,00160	0,00076	0,00203	0,00097	0,00175	0,00081
rsa(r)	1,03480	3,76911	0,90324	3,42288	1,05136	3,77224	1,01198	3,64629
	-0,00443	-0,00718	-0,00287	-0,00652	-0,0045	-0,00719	-0,00433	-0,00311
mse(r)	0,00243	0,00364	0,00285	0,00403	0,00255	0,00373	0,00290	0,00413

Part c. $k = 100$

Item	Form1		BMDBPBak		S-PLUS		MINITAB	
	$p = 0.10$	$p = 0.50$	$p = 0.10$	$p = 0.50$	$p = 0.10$	$p = 0.50$	$p = 0.10$	$p = 0.50$
	$n = 100$	$n = 500$	$n = 100$	$n = 500$	$n = 100$	$n = 500$	$n = 100$	$n = 500$
m(r) = m	0,49708	0,49669	0,49790	0,49753	0,49724	0,49686	0,49772	0,49739
bi(r)	-0,00016	-0,00055	0,00066	0,00029	0,00000	-0,00038	0,00048	0,00015
	-0,00029	-0,00017	-0,00024	-0,00015	-0,00028	-0,00017	-0,00027	-0,00016
rbi(r)	-0,00032	-0,00111	0,00113	0,00057	0,00001	-0,00076	0,00097	0,00030
	-0,00058	-0,00034	-0,00048	-0,0003	-0,00057	-0,00033	-0,00054	-0,00032
s(r)	0,00289	0,00378	0,00238	0,00335	0,00284	0,00370	0,00268	0,00354
	-0,00021	-0,00012	-0,00017	-0,00011	-0,0002	-0,00012	-0,00019	-0,00011
asm(r)	0,00209	0,00104	0,00142	0,00072	0,00191	0,00094	0,00156	0,00078
rsa(r)	1,14333	3,34737	0,94122	2,96039	1,12333	3,27483	1,05984	3,13142
	-0,00489	-0,00638	-0,00403	-0,00564	-0,00481	-0,00624	-0,00454	-0,00267
mse(r)	0,00273	0,00323	0,00304	0,00363	0,00284	0,00332	0,00316	0,00369

Table 7: Comparison of simulation results with the MA(1) model with contaminations of p = 0.10 and 0.50, n = 100 and 500. Parameter $\alpha = 0.90$ $(\rho = 0.49724), T = 100$, *and 3 values of k*

Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., and Tukey, J. W. (1972). *Robust estimates of location: survey and advances*. Princeton University Press, Princeton.

BMDP (1990). *BMDP Statistical Software*. Los Angeles, California.

BOX, G. E. P. and JENKINS, G. M. (1970). *Time Series Analysis Forecasting and Control*. Holden-Day, San Francisco.

BOX, G. E. P. and TIAO, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*, 70:70–79.

BROCKWELL, P. J. and DAVIS, R. A. (1991). *ITSM: An Interactive Time Series Modelling Package for the PC*. Springer-Verlag, New York.

CHANG, I., TIAO, G. C., and CHAN, C. (1988). Estimation of Time Series parameters in the presence of outliers. *Technometrics*, 30:193–204.

COOK, R. D. and WEISBERG, S. (1982). *Residuals and influence in regression*. Chapman and Hall, London.

FOX, A. J. (1972). Outliers in Time Series. *Journal of the Royal Statistical Society*, B 43:350–363.

LINDSAY, B. G. (1995). Mixture models: theory, geometry and applications. Regional conference series in probability and statistics, Institute of Mathematical Statistics.

MARTIN, R. D. (1980). Robust estimation of autoregressive models. Directions in time series, Institute of Mathematical Statistics, Hayward, California.

MINITAB (1996). *Guía del Usuario de MINITAB en Español, Versión 2 para Windows*. Minitab Inc., Los Angeles, California, US.

PEÑA, D. (1990). Influential observations in Time Series. *Journal of Business and Economics Statistics*, 8:235–241.

TIKKU, M. L., WONG, W. K., VAUGHAN, D. C., and BIAN, G. (2000). Time Series models in non-normal situations: symmetric innovations. *Journal of Time Series Analysis*, 21:571–596.

VENABLES, W. N. and RIPLEY, B. D. (1997). *Modern Applied Statistics with S-Plus*. Springer-Verlag, New York, 2nd ed.

WOLFRAM, S. (1991). *Mathematica*. Addison-Wesley Publishing Co., Reading, Massachusetts, US, 2nd ed.