Robust Simulation-Based Estimation of ARMA Models

Xavier DE LUNA and Marc G. GENTON

This article proposes a new approach to the robust estimation of a mixed autoregressive and moving average (ARMA) model. It is based on the indirect inference method that originally was proposed for models with an intractable likelihood function. The estimation algorithm proposed is based on an auxiliary autoregressive representation whose parameters are first estimated on the observed time series and then on data simulated from the ARMA model. To simulate data the parameters of the ARMA model have to be set. By varying these we can minimize a distance between the simulation-based and the observation-based auxiliary estimate. The argument of the minimum yields then an estimator for the parameterization of the ARMA model. This simulation-based estimation procedure inherits the properties of the auxiliary model estimator. For instance, robustness is achieved with GM estimators. An essential feature of the introduced estimator, compared to existing robust estimators for ARMA models, is its theoretical tractability that allows us to show consistency and asymptotic normality. Moreover, it is possible to characterize the influence function and the breakdown point of the estimator. In a small sample Monte Carlo study it is found that the new estimator performs fairly well when compared with existing procedures. Furthermore, with two real examples, we also compare the proposed inferential method with two different approaches based on outliers detection.

Key Words: Breakdown point; Indirect inference; Influence function; GM-estimator; Robustness; Time series.

1. INTRODUCTION

The concept of robustness in statistics is usually defined as the lack of sensitivity of a particular inferential procedure to departures from the model assumptions. A theory of robustness (Huber 1981; Hampel, Ronchetti, Rousseeuw, and Stahel 1986) has developed for departures arising from contaminated observations—called outliers—present in a "clean" dataset; that is, generated from a known model. Typically maximum likelihood and least squares estimators have poor robustness properties and several alternative robust estimators have been proposed, such as M estimators (Huber 1981) and L_1-based estimators. Robust

Xavier de Luna is Research Fellow, Department of Economics, Umeå University, S-90187 Umeå (E-mail: xavier.deluna@econ.umu.se). Marc G. Genton is Assistant Professor, Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203 (E-mail: genton@stat.ncsu.edu).

©2001 *American Statistical Association, Institute of Mathematical Statistics, and Interface Foundation of North America*

Journal of Computational and Graphical Statistics, Volume 10, Number 2, Pages 370–387

estimation in the time series context is a difficult task because different types of outliers may occur. For instance, outliers can replace, or be added to, some observations of the stochastic process. They can also be found in the innovations driving the process. Furthermore, the configuration of time points where the contaminations occur becomes important: isolated and patchy outliers can have different effects and both have been observed in practice. In the last two decades, the subject of robust estimation for time series models has developed quickly; see Martin and Yohai (1985) for a survey. The earlier works put particular emphasis on autoregressive (AR) processes. In this case, M estimators were shown to be nonrobust, and were generalized to GM estimators in order to obtain the desired robustness properties (Künsch 1984). Estimators based on residual autocovariances, RA estimators, were later proposed by Bustos and Yohai (1986) as alternatives. Unfortunately, GM and RA estimators are not robust for moving average or mixed autoregressive moving average (ARMA) models, since a single outlier at a certain time influences all subsequent terms in the estimation equations. Several ad hoc methods have been proposed to overcome this difficulty. For example, estimators based on truncated residual autocovariances (TRA estimators) were also introduced by Bustos and Yohai (1986). More recently, Allende and Heiler (1992) have proposed the use of recursive GM estimators, denoted RGM estimators. They performed a Monte Carlo study, showing that their proposal outperforms all other existing methods. Unfortunately, due to the complexity of the algorithm, neither an asymptotic distribution nor theoretical robustness results are available to date. Moreover, computation of RGM estimators is costly because it involves several GM estimations. This article proposes a robust estimation method based on simulated data and on a single GM estimation. An essential feature of the introduced estimator, compared to existing robust estimators for ARMA models, is its theoretical tractability that allows us to show consistency and asymptotic normality, as well as to characterize the corresponding influence function and the breakdown point.

This novel robust estimator is inspired from the indirect inference method introduced by Gouriéroux, Monfort, and Renault (1993). Indirect inference is part of an increasingly growing tool-kit of simulation-based inference procedures. These are often proposed with the purpose of dealing with complex models whose likelihood function is intractable; Diggle and Gratton (1984) is an early reference. The indirect inference allows for the estimation of the parameter, say θ, of a model of interest whose likelihood or any other classic estimating criterion is not available neither in closed form nor algorithmically. The indirect inference consists in using an auxiliary model with parameterization π, whose estimation is feasible. This auxiliary model is fitted to the observed data, yielding $\hat{\pi}$, as well as to data simulated from the model of interest with a given value of θ, giving π^*. Then, $\hat{\pi}$ and π^* are calibrated by choosing a value for θ, the final estimate.

We borrow the indirect inference estimation algorithm and modify it in order to obtain a robust device for the adjustment of ARMA models. The idea is to use an autoregressive representation of an ARMA process as an auxiliary model. Indeed, while existing robust inference methods for ARMA models are not satisfactory because of their essentially ad hoc nature, the robust GM estimation of AR models is well mastered both from a theoretical and a practical point of view.

This article is arranged as follows. Section 2 presents different contamination models for ARMA processes, and Section 3 describes in detail the robust simulation-based estimation

algorithm. Section 4 analyzes the asymptotic properties of our method. In particular, known asymptotic results for the GM estimation of AR models are readily "transmitted" to the simulation-based estimator for the ARMA parameters. Section 5 demonstrates a Monte Carlo experiment, where small sample efficiency and robustness are studied and compared with existing estimators. Section 6 presents two case studies. Section 7 summarizes our main conclusions.

2. CONTAMINATION MODELS

Let $\{X_t, t \in \mathbb{Z}\}$ be a stationary stochastic process. We call it an autoregressive and moving average process of order p and q, ARMA(p, q), if it admits the representation

$$X_t - \mu = \sum_{i=1}^{p} \alpha_i (X_{t-i} - \mu) + \sum_{j=1}^{q} \beta_j Z_{t-j} + Z_t. \tag{2.1}$$

The so called innovation process $\{Z_t\}$ is assumed to be independently and identically distributed with mean zero and finite variance σ^2.

Subsequently, we assume that $\alpha_1, \ldots, \alpha_p$ and $\beta_1, \ldots, \beta_q$ are such that the process is causal and invertible. Brockwell and Davis (1991) is a general reference on causal and invertible ARMA processes. Note that $\{X_t\}$ is called autoregressive process when $q = 0$ and moving average process when $p = 0$.

As noted in Section 1, robustness is typically more difficult with dependent than with independent data, because in the former there is a need to consider several different possible types of contamination. The three main types of outliers generally considered for dependent data are: innovation outliers (IO), which affect all subsequent observations, additive outliers (AO), and replacement outliers (RO), which have both no effect on subsequent observations.

The ARMA(p, q) process $\{X_t, t \in \mathbb{Z}\}$ is said to have innovation outliers (IO) if it satisfies (2.1) and the innovation process $\{Z_t\}$ has a heavy-tailed distribution; for instance, $F_\varepsilon = (1 - \varepsilon)F + \varepsilon H$, where ε is small and H is an arbitrary distribution with greater dispersion than F. The important characteristic of this kind of outliers is that the contaminated time series comes from a perfectly observed ARMA(p, q) process. For that reason, naive robust estimators, such as M estimators, and even maximum likelihood estimators can typically cope with IO.

The process $\{Y_t, t \in \mathbb{Z}\}$ is said to have additive outliers (AO) when defined by $Y_t = X_t + B_t W_t$, where $\{X_t\}$ is an ARMA(p, q) process satisfying (2.1), $\{B_t\}$ is a Bernoulli process with $P(B_t = 1) = \varepsilon$, and $\{W_t\}$ is an independent sequence of random variables, independent of the sequences $\{X_t\}$ and $\{B_t\}$. The process $\{X_t\}$ is observed with probability $1 - \varepsilon$, whereas the process $\{X_t\}$ plus an error $\{W_t\}$ is observed with probability ε. $\{Y_t\}$ is therefore not an ARMA process. AO are known to be much more dangerous than IO. Note also that additive outliers have a similar effect to replacement outliers (RO), where the contaminated process is defined as $Y_t = (1 - B_t)X_t + B_t W_t$. This means that the process $\{X_t\}$ is observed with probability $1 - \varepsilon$, and replaced by an error $\{W_t\}$ with probability ε. Here again, $\{Y_t\}$ does not admit representation (2.1). Isolated or patchy outliers can be modeled with respectively an independent or strongly dependent process $\{B_t\}$.

3. ESTIMATION METHOD

3.1 SIMULATION-BASED ESTIMATION ALGORITHM

Maximum likelihood estimators and alikes are known to be very sensitive to the presence of AO or RO. We therefore propose an alternative estimation method of the parameter $\boldsymbol{\theta} = (\mu, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q, \sigma^2)'$, which uses the auxiliary representation of the stochastic process $\{X_t\}$:

$$X_t - \mu = \sum_{i=1}^{r} \pi_i (X_{t-i} - \mu) + U_t, \tag{3.1}$$

where π_i, $i = 1, \ldots, r$, are defined such that $E(U_t \mathbf{V}_t) = \mathbf{0}$, with $\mathbf{V}_t = (X_{t-1} - \mu, \ldots, X_{t-r} - \mu)'$. Then, $\{U_t\}$ is a process with mean zero and variance σ_r^2. We call $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_r, \sigma_r^2)'$ the auxiliary parameter. Note that the parameter $\boldsymbol{\pi}$ is solution of $r + 1$ Yule–Walker equations; see, for example, Brockwell and Davis (1991, sec. 8.1). For instance, if the process is ARMA(0,1) and for $r = 1$, we have $\pi_1 = \beta_1/(1 + \beta_1^2)$ and $\sigma_1^2 = \sigma^2(1 + \beta_1^2)(1 - \pi_1^2)$. We assume in the sequel that the nontrivial case $q > 0$ applies. In the case $q = 0$, the robust estimation of $\boldsymbol{\theta}$ can be done with GM estimators as described in Section 3.2. Finally note that, because $q > 0$, $\{U_t\}$ is a correlated process, although U_t is uncorrelated with $\mathbf{V}_t$ by definition.

Remark 1. In this article, we talk about auxiliary autoregressive representation AR(r) when referring to (3.1) and about autoregressive process when referring to (2.1) with $q = 0$. In particular, when $q > 0$, $\{X_t\}$ admits an auxiliary AR(r) representation although it is not an autoregressive process.

Assume that the realization $\mathbf{y} = (y_1, \ldots, y_n)'$ of $\{Y_t\}$, a possibly contaminated ARMA(p, q) stochastic process, is observed. The estimator of $\boldsymbol{\theta}$ proposed in this article is based on the following algorithm:

Step 0: Estimate μ with $\hat{\mu}$ such that

$$\sum_{t=1}^{n} \psi(y_t - \hat{\mu}) = 0, \tag{3.2}$$

where $\psi(\cdot)$ is a real valued function.

Step 1: Estimate the auxiliary parameter $\boldsymbol{\pi}$ with $\hat{\boldsymbol{\pi}} = (\hat{\pi}_1, \ldots, \hat{\pi}_r, \hat{\sigma}_r^2)'$ such that

$$\sum_{t=r+1}^{n} \eta(\mathbf{v}_t, u_t/\hat{\sigma}_r)\mathbf{v}_t = \mathbf{0}, \tag{3.3}$$

$$\sum_{t=r+1}^{n} \chi(u_t/\hat{\sigma}_r) = 0, \tag{3.4}$$

where $\mathbf{v}_t = (y_{t-1} - \hat{\mu}, \ldots, y_{t-r} - \hat{\mu})'$, $u_t = (y_t - \hat{\mu}) - \hat{\pi}_1(y_{t-1} - \hat{\mu}) - \cdots - \hat{\pi}_r(y_{t-r} - \hat{\mu})$ and $\eta(\cdot)$ and $\chi(\cdot)$ are suitable real valued functions.

Step 2: Simulate $m = sn$ independent pseudo-realizations from the standard normal distribution, where s is a strictly positive integer.

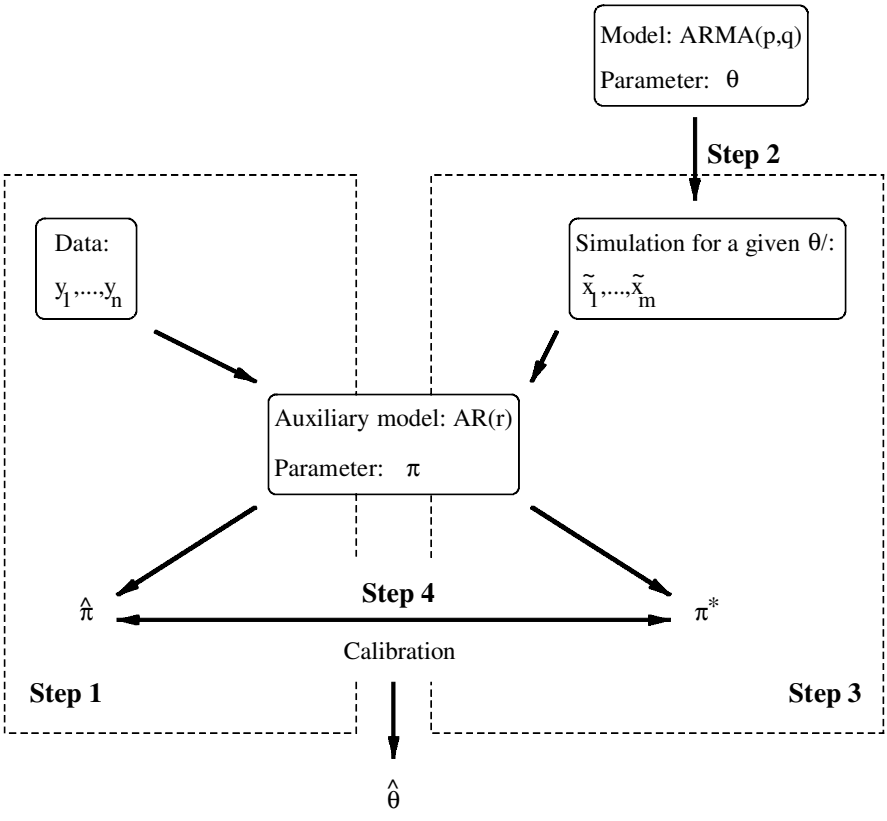

Figure 1. Schematic Illustration of the Simulation-Based Estimation Algorithm.

Step 3: Use the simulated random sample in Step 2 as innovations and a given value of $\boldsymbol{\theta}$ (with $\mu = 0$) to generate a sequence $\tilde{\mathbf{x}} = (\tilde{x}_1, \ldots, \tilde{x}_m)$ from model (2.1). Estimate $\boldsymbol{\pi}$ with an efficient (when possible) estimator denoted $\boldsymbol{\pi}^*$; when innovations are normally distributed, maximum likelihood, least squares, and Yule–Walker estimators are asymptotically equivalent, see, for example, Brockwell and Davis (1991).

Step 4: The estimator $\hat{\boldsymbol{\theta}}$ of the parameter $\boldsymbol{\theta}$ is obtained by minimizing the quadratic distance between $\boldsymbol{\pi}^*$, which is a function of $\boldsymbol{\theta}$ by construction, and $\hat{\boldsymbol{\pi}}$:

$$(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}^*)' \Omega (\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}^*), \tag{3.5}$$

where Ω is a weighting matrix, whose optimal choice is discussed later. Note that criterion (3.5) is minimized by letting $\boldsymbol{\theta}$ vary in Step 3 while the simulated innovations in Step 2 are kept fixed throughout. The latter point is important to ensure the success of the estimation algorithm; see Gouriéroux, Monfort, and Renault (1993).

A schematic illustration of the algorithm is given in Figure 1.

When the estimator $\hat{\boldsymbol{\pi}}$ and $\boldsymbol{\pi}^*$ are based on the same criterion—for example, both are least squares estimators—we obtain the indirect inference estimator introduced by Gouriéroux et al. (1993). We obtain a robust estimator $\hat{\boldsymbol{\theta}}$ by using a robust estimation procedure in Steps 0 and 1. On the other hand, in Step 3 no robustification is needed be-

cause the data are simulated and therefore outliers-free. In this article we study the use of GM estimation in Step 1.

Remark 2. Note that the simulation-based algorithm has been described for normally distributed innovations (Step 2) for convenience of exposition. The normal distribution is also the most natural choice and corresponds to the widespread use of the Gaussian likelihood. However, while maximum likelihood estimation is generally intractable for non-Gaussian processes, the simulation-based estimation algorithm is easily adaptable to other innovations distributions as long as they allow for the generation of pseudo-realizations in Step 2. The choice of an adequate distribution should as usual be based on scientific knowledge of the problem at hand and/or on data exploration.

3.2 GM ESTIMATION OF AN AR MODEL

The GM estimator for the parameter of an AR model has been described and justified extensively in the literature; see, for example, Denby and Martin (1979), Martin (1980), Bustos (1982), Künsch (1984), and Bustos and Yohai (1986). Therefore, we here only succinctly describe the GM estimator and refer the interested reader to the above-mentioned literature. For Step 0 and 1 to be robust against the AO and RO types of outliers, the functions $\psi(\cdot)$ and $\chi(\cdot)$ must be bounded and η must be such that $\eta(\mathbf{v}, u)\mathbf{v}$ is bounded. This is not the case, for instance, for the least squares estimator, where $\eta(\mathbf{v}, u) = u$, or classical M estimators, where $\eta(\mathbf{v}, u) = \eta(u)$. Several proposals for η of the form $\eta(\mathbf{v}, u) = w_1(\mathbf{v})\psi(w_2(\mathbf{v})u)$, for appropriate functions $\psi : \mathbb{R} \to \mathbb{R}$ and weight functions $w_1, w_2 : \mathbb{R}^r \to \mathbb{R}_+$, can be found in the literature. The choice of functions ψ, χ and η are further restricted by the consistency requirement, see Section 4.3. There are two classic choices for η. The first defines the Mallows GM estimator (Mallows 1976), and is such that $w_1(\mathbf{x}) = \psi_1(\|\mathbf{x}\|)/\|\mathbf{x}\|$ and $w_2(\mathbf{x}) = 1$, for an appropriate norm $\|\cdot\|$. The second is the Hampel–Krasker–Welsch GM estimator defined with $w_1(\mathbf{x}) = 1/\|\mathbf{x}\|$ and $w_2(\mathbf{x}) = \|\mathbf{x}\|$, see, for example, Martin and Yohai (1985). More details can also be found in Section 5.1 where particular examples of GM estimators are described and used.

4. ASYMPTOTIC THEORY

4.1 PRELIMINARIES

Without loss of generality we assume that $\mu = 0$ in the following theoretical developments. We need to define the function $h : \mathcal{M} \subset \mathbb{R}^{p+q+1} \to \mathcal{A} \subset \mathbb{R}^{r+1}$ such that $h(\boldsymbol{\theta}) = \boldsymbol{\pi}$ for $\boldsymbol{\theta} \in \mathcal{M}$ and $\boldsymbol{\pi} \in \mathcal{A}$; that is, the function binding the parameterization of representation (3.1) and model (2.1). The subspaces $\mathcal{M}$ and $\mathcal{A}$ are defined by the causality and invertibility assumptions on model (2.1). Standard asymptotic results for the estimation algorithm proposed in Section 3.1 can be deduced when the function $h(\cdot)$ is locally injective around $\boldsymbol{\pi}_0 = h(\boldsymbol{\theta}_0)$, where $\boldsymbol{\theta}_0$ denotes the unknown unique value for $\boldsymbol{\theta}$ for which model (2.1) holds,

and the matrix of the first partial derivatives (Jacobian matrix) $D(\boldsymbol{\theta}_0) = \partial h(\boldsymbol{\theta})/\partial \boldsymbol{\theta}'|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ exists. The first property holds when $r \geq p + q$. Note that there is a one-to-one mapping between $\mathcal{A}$ and $\mathcal{G} \subset \mathbb{R}^{r+1}$ the space defined by the vector $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_r)'$ of autocovariances up to order r of the causal ARMA process $\{X_t\}$. This mapping is actually defined by the Yule–Walker equations. These equations remain valid for (3.1). The function h is then injective when $r \geq p+q$ because there is an injective mapping between $\mathcal{M}$ and $\mathcal{G}$. This latter assertion is a consequence of the equations linking $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$ given by Brockwell and Davis (1991, p. 93). We, therefore, assume in the sequel that $r \geq p+q$. Finally, the existence of the Jacobian matrix $D(\boldsymbol{\theta}_0)$ is a straightforward consequence of $\boldsymbol{\pi}$ being differentiable with respect to $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}$ being differentiable with respect to $\boldsymbol{\theta}$.

Note that when $p = 0$, the function $\boldsymbol{\pi} = h(\boldsymbol{\theta})$ has a known analytical form; as an example we gave the function for the ARMA(0,1) case in the text after Equation (3.1). However, when $p > 0$, only an algorithmic characterization of the function $h(\cdot)$ is available (Brockwell and Davis 1991, p. 93). The main consequence is that the inverse of the function $h(\cdot)$ is not retrievable in general, and an estimate of $\boldsymbol{\theta}$ cannot be constructed as $h^{-1}(\hat{\boldsymbol{\pi}})$, thereby justifying the use of the simulation step in our proposal.

4.2 Influence Function and Breakdown Point

The influence function (Hampel 1974) is a tool to describe the robustness properties of an estimator. Its importance lies in its appealing heuristic interpretation: it measures the asymptotic bias caused by an infinitesimal contamination of the observations. Hampel's original definition was aimed at the independent and identically distributed case. Its extension to the time series setting has given raise to two different approaches proposed by Künsch (1984) and Martin and Yohai (1986).

Künsch's (1984) influence function is bounded when the GM estimation of the auxiliary AR representation is performed and X_t is a causal and invertible ARMA process. Martin and Yohai's (1986) definition requires the η function to be based on a redescending ψ to ensure boundedness of the influence function. This is not the case when using our simulation-based estimation algorithm.

Both Künsch's and Martin and Yohai's influence functions can be derived for the simulation-based estimator of $\boldsymbol{\theta}$ from the influence function of the estimator of $\boldsymbol{\pi}$. Denote by Π and Θ the statistical functionals corresponding respectively to the estimators $\hat{\boldsymbol{\pi}}$ defined by Equations (3.3) and (3.4), and $\hat{\boldsymbol{\theta}}$ defined by Equation (3.5). Let $\mathrm{IF}_{\mathrm{AR}}(\nu, \Pi, F)$ be the vector influence function of the estimator of $\boldsymbol{\pi}$, and $\mathrm{IF}_{\mathrm{ARMA}}(\nu, \Theta, F)$ be the vector influence function of the simulation-based estimator of $\boldsymbol{\theta}$. Here F is the $(r+1)$-dimensional marginal of the joint distribution of the process, and ν is the probability measure defining the contamination process. The function h defined in the previous section sends Θ on $\Pi = h(\Theta)$. The relation between the two influence functions is (Genton and de Luna 2000, theorem 1):

$$\mathrm{IF}_{\mathrm{ARMA}}(\nu, \Theta, F) = P(\Theta(F))\, \mathrm{IF}_{\mathrm{AR}}(\nu, \Pi, F), \tag{4.1}$$

where $P(\Theta(F)) = \left[D(\Theta(F))'\Omega D(\Theta(F))\right]^{-1} D(\Theta(F))'\Omega$. Thus, these two influence functions are proportional so that the boundedness of $\mathrm{IF}_{\mathrm{AR}}$ implies the boundedness of $\mathrm{IF}_{\mathrm{ARMA}}$.

The breakdown point is another important feature of reliability of an estimator (Huber 1981, 1984; Hampel et al. 1986). It indicates how many data points need to be replaced by arbitrary values to destroy the estimator, i.e. to bring the estimator to the boundaries of the parameter space. This concept can be applied to a dependent data setting; see, for example, Genton (1998) for a spatial random field situation. It is important to note, however, that the effect of a fraction of outliers may be different, depending upon whether they occur in isolation or in patches (Ma and Genton 2000). For AR models, it turns out that least squares and M estimators both have zero breakdown point. Martin and Jong (1977), Martin (1980), and Martin and Yohai (1985) studied the breakdown point of GM estimators for AR models. They argue that in this case, the breakdown point is positive, but bounded above by $1/(r+1)$, where r is the order of the autoregression. The simulation-based estimator of $\boldsymbol{\theta}$ has the same positive breakdown point, bounded above by $1/(r+1)$. Once again, the reason is that in Step 3 of our algorithm, the simulated data are outliers-free.

Finally, one can expect that other robustness properties of GM estimators for the auxiliary autoregressive representation will be transmitted to the simulation-based estimator of $\boldsymbol{\theta}$. For instance, the concept of qualitative robustness has been generalized to the time series context by Boente, Fraiman, and Yohai (1987), by defining weak and strong resistance. In particular, they showed the strong resistance of GM estimators for autoregressive models.

4.3 CONSISTENCY AND ASYMPTOTIC NORMALITY

Bustos (1982) showed for a fairly wide class of ergodic processes—including causal and invertible ARMAs—that the GM estimation of $\boldsymbol{\pi}$, defined by (3.3) and (3.4), is consistent and asymptotically normal. We thus have, when n tends to infinity,

$$\sqrt{n}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}_0) \sim N(\mathbf{0}, V^{\#}), \tag{4.2}$$

where $V^{\#}$ can be found in Bustos (1982, theorem 2.2). For this asymptotic result to hold, regularity conditions on the functions η and χ, in (3.3) and (3.4) respectively, need to be imposed; see again Bustos (1982) for more details.

Because the binding function $h(\boldsymbol{\theta})$ is injective, the estimator $\hat{\boldsymbol{\theta}}$ defined in Section 3.1 inherits the consistency and asymptotic normality from $\hat{\boldsymbol{\pi}}$. More specifically, when (4.2) holds, for a given matrix Ω and s fixed,

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \sim N(\mathbf{0}, V), \tag{4.3}$$

when n tends to infinity. Here

$$V = P(\boldsymbol{\theta}_0) V^{\#} P(\boldsymbol{\theta}_0)' + \frac{1}{s} P(\boldsymbol{\theta}_0) V^{*} P(\boldsymbol{\theta}_0)', \tag{4.4}$$

where V^* is the asymptotic variance of $\boldsymbol{\pi}^*$. Result (4.3) is a direct consequence of Proposition 4.2 in Gouriéroux and Monfort (1996).

Note that $\hat{\boldsymbol{\theta}}$ is consistent for any $s > 0$ while its efficiency is improved when increasing s. For s large enough, we obtain the best efficiency with $\Omega = V^{\#-1}$. The matrix V simplifies then to

$$V \cong \left[D(\boldsymbol{\theta}_0)' V^{\#-1} D(\boldsymbol{\theta}_0) \right]^{-1}.$$

Remark 3. In practice, $\Omega = V^{\#-1}$ has to be expressed as a function of $\boldsymbol{\theta}$ or need to be estimated. However, choosing $\Omega = Id$ simplifies the computation effort while, from practical experience, the loss of efficiency is often limited. For those cases where the matrix $V^{\#-1}$ is far from being diagonal—strong correlation structure in the components of the sum of squares (3.5) to be minimized—it is sometimes possible and convenient to use a slight modification of (3.5) to improve on the efficiency; see next section for details.

Remark 4. A consistent estimator for V can be obtained by replacing $D(\boldsymbol{\theta}_0)$ by a numerical approximation of $\partial\pi^*(\boldsymbol{\theta})/\partial\boldsymbol{\theta}'|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$ (see Genton and de Luna 2000), and using a consistent estimator for $V^{\#}$, see Gouriéroux et al. (1993, appendix 2).

Remark 5. Asymptotic properties have been derived for any value of $r \geq p + q$ fixed. Increasing r may improve the efficiency if there is structure left in the process U_t (noise of the auxiliary autoregression) although only marginally after a given level. Thus, a heuristic approach to the choice of r is to use an information criterion such as the Akaike information criterion (AIC) or the Bayesian information criterion (BIC); see, for example, de Luna (1998), ensuring thus that as little structure as possible remains in the process U_t. One may then repeat the procedure for a few other values of r to ensure that the estimation is not sensitive to an increase in r. In fact the efficiency performances of the estimator will in general be similar for a fairly wide range of values of r, see Section 5.2.

5. MONTE CARLO STUDY

5.1 COMPUTATION

The computation of GM estimators (Step 1) can be done with an iterative weighted least squares algorithm; see Martin (1980). This is implemented in the widely spread statistical package S-Plus; the built-in command is `ar.gm`.

In the Monte Carlo study reported in this section, we have used the minimization routine E04FDF (Gauss–Newton) from the Fortran NAG library to perform Step 4. The starting values were chosen as the true values for the simulation study and as the ML estimates for the illustrative examples of Section 6. Pseudo-random numbers for Step 2 were obtained via the algorithm of linear congruences; see Rubinstein (1981). The Durbin–Levinson algorithm (see Brockwell and Davis 1991) solving the Yule–Walker equations is used in Step 3. Here maximum likelihood and least squares could also have been used without implying significant change for large samples. The GM estimation was performed with the `ar.gm` command. The latter is based on the Mallows type η function and allows the user to choose between the two following ψ functions: the Huber family $\psi_H(x) = \min\{|x|, c\}\text{sign}(x)$ and the bi-square family $\psi_B(x) = x(1 - (x/c)^2)^2$, for $0 \leq x \leq c$, zero otherwise. Here c is a tuning parameter which regulates the trade-off between robustness and efficiency under the noncontaminated model. The command `ar.gm` allows us to start the iterations of the weighted least squares algorithm with $\psi_H(x)$ (Equation (3.3) has then a unique solution) and to finish with $\psi_B(x)$. Two strategies are adopted in this study. The first uses three iterations

with ψ_B (denoted IGM-a in the sequel). It corresponds to setting the `ar.gm` parameters `iterh=0` and `iterb=3`. The second estimator (IGM-b) is obtained with four iterations with ψ_H and one with ψ_B, corresponding to `iterh=4` and `iterb=1`. The values for c are automatically set once the user has chosen the value of the parameter `effgm`, controlling the efficiency, in `ar.gm`; we have used throughout `effgm=0.7`. For the innovations scale estimation, Huber's proposal 2 is used: $\chi(x) = \psi^2(x) - E_\Phi[\psi^2]$, where Φ is the standard normal distribution.

For ARMA(p, q) processes with $p, q > 0$, we use a computational modification of the algorithm described in Section 3. For the clarity of the exposition we specialize here to the ARMA(1,1) with zero mean process: $X_t - \alpha X_{t-1} = \beta Z_{t-1} + Z_t$. Note that $\{D_t\} = \{X_t - \alpha X_{t-1}\}$ is a moving average process. We use this fact to simplify the computation. In Step 3 we can simulate a sample path from $\{D_t\}$ rather than $\{X_t\}$, and then fit the auxiliary AR representation

$$D_t = \sum_{i=1}^{r-1} \tilde{\pi}_i D_{t-i} + U_t. \tag{5.1}$$

yielding the estimator $\tilde{\boldsymbol{\pi}}^* = (\tilde{\pi}_1^*, \ldots, \tilde{\pi}_{r-1}^*, \sigma_r^{2*})'$. Comparing (5.1) with (3.1) we see that $\pi_1 = \alpha + \tilde{\pi}_1, \pi_2 = \tilde{\pi}_2 - \alpha\tilde{\pi}_1, \ldots, \pi_{r-1} = \tilde{\pi}_{r-1} - \alpha\tilde{\pi}_{r-2}$, and $\pi_r = -\alpha\tilde{\pi}_{r-1}$. Hence, $\boldsymbol{\pi}^*$ in (3.5) is computed from $\tilde{\boldsymbol{\pi}}^*$. This manner of proceeding has several advantages. A moving average and not a more complex mixed autoregressive and moving average sample path has to be simulated. From our experience the minimization of (3.5) is faster to converge with the above modification. Moreover, when the components of $\hat{\boldsymbol{\pi}}$ are strongly correlated the computation of a weighting matrix (see Remark 4) can be spared with the use of $\tilde{\pi}_1^*, \ldots, \tilde{\pi}_{r-1}^*$ as some kind of instrumental variables; that is, calibrating $\hat{\boldsymbol{\pi}}$ with $(\alpha + \tilde{\pi}_1^*, (\tilde{\pi}_2^* - \alpha\tilde{\pi}_1^*)\tilde{\pi}_1^*, \ldots, (\tilde{\pi}_{r-1}^* - \alpha\tilde{\pi}_{r-2}^*)\tilde{\pi}_{r-2}^*, -\alpha\tilde{\pi}_{r-1}^{*2}, \sigma_r^{2*})'$ instead of $\boldsymbol{\pi}^*$ can dampen the correlation structure of the components of the sum of squares (3.5) to be minimized, without adding much extra computational burden. This modified Step 3 together with the instruments are used in the experiments described below concerning the ARMA(1,1) processes.

The computational time to obtain a simulation-based estimate is reasonable since the actual simulation of random numbers is performed only one time (Step 2 of the algorithm). For the models and sample sizes considered in the next sections, the computational time to obtain the estimator $\hat{\boldsymbol{\theta}}$ varied typically between less than one second and 40 seconds on a Silicon Graphics computer using a 200 MHZ processor. The variation was mainly due to slower or faster convergence when minimizing (3.5).

5.2 DESIGN OF THE STUDY

We have chosen to perform a Monte Carlo experiment which allows for comparison with Allende and Heiler's (1992) published results. We therefore follow partially their design but consider also other situations such as different types of outliers. Thus, the study is based on three ARMA(0,1) processes, $X_t = \beta Z_{t-1} + Z_t$, with $\beta = -0.5, -0.8$, and 0.5, and one ARMA(1,1), $X_t = \alpha X_{t-1} + \beta Z_{t-1} + Z_t$, with $\alpha = 0.8$ and $\beta = 0.5$. The innovations are simulated from the normal distribution with variance $\sigma^2 = 1$ in all cases. To study the robustness properties of different estimation methods we consider AO processes,

$Y_t = X_t + B_t W_t$, and RO processes, $Y_t = (1 - B_t)X_t + B_t W_t$ (as defined in Section 2), with $P(B_t = 1) = \varepsilon = 0.05$, $\text{var}(W_t) = \tau^2 \text{var}(X_t)$ with $\tau^2 = 9,100$. Time series of length $n = 100$ were generated with S-Plus. The estimation methods studied are the maximum likelihood (ML)—the S-Plus built-in command is used—and three versions of the simulation-based algorithm of Section 3: the first (denoted ILS for indirect least squares) uses least squares in Step 1; the two others are IGM-a and IGM-b as described previously. The simulation-based algorithm was used with $s = 30$ for ARMA(0,1) cases and $s = 90$ for ARMA(1,1) cases. While consistency is independent of s, efficiency is improved by increasing its value. Increasing s over the chosen values did not, however, bring significant efficiency improvement in the experiments reported here. Another tuning parameter of the algorithm is r, the order of auxiliary AR representation. It needs to be large enough for the representation (3.1) to be a good approximation of (2.1), depending on the values of the parameters α and β. We have used the following values: $r = 5$ when $|\beta| = 0.5$ and $r = 10$ for $\beta = -0.8$, adding one when $p = 1$. From nonreported experiments, we have noticed that too large values for r do not have much influence on the efficiency of the estimators under the uncontaminated model but weakens the robustness properties.

5.3 Results and Comments

Results are displayed in Tables 1–3. The tables contain the average of the 500 estimators obtained for α and β, the corresponding mean squared errors (MSEs) as well as relative efficiencies (eff), defined as the ratio of the MSE of the ML estimator to the MSE of the estimator under consideration. Where applicable Allende and Heiler's (1992) results are put in parallel. We report only the efficiencies they obtained with RGM, because the latter estimator was the best performer; in particular, RGM estimators were superior to RA and TRA estimators. Because they consider eight different RGM estimators Tables 1–3 give, in a row labeled RGM, the worst and best efficiencies Allende and Heiler (1992) obtained. These can be directly compared with our own efficiency results.

In the up-coming comments we use the words "efficient" and "robust" for the same concept of performance; that is, an efficient (or robust) estimator is called so when its sample MSE is low (always in comparison with one or several other estimator). We, however, systematically talk about efficiency in the noncontaminated situations while robustness always refers to the presence of outliers, either AO or RO.

Table 1 and 2 report the results of our experiments concerning the purely moving average models. We first comment on the difference between ML and ILS estimation which is of interest per se. Both are nonrobust with ML being asymptotically efficient. ILS was advocated by Gouriéroux et al. (1993) as an alternative to ML. They showed that the ILS estimator, which is faster to compute, can have similar small sample efficiency to ML. These results are confirmed here by our experiments although for a positive β, here 0.5 (see Table 1), the loss of efficiency when using ILS is not negligible. Finally, it is apparent when looking at biases and MSEs that both ML and ILS are not able to cope with contaminated situations, neither AO nor RO.

When replacing least squares with GM estimation in Step 2 we gain a great deal in robustness when both AO or RO are present as can be seen in Tables 1 and 2. Both versions

Table 1. Monte Carlo Results for the ARMA(0,1) Case With $\beta = -0.5$

	Clean data			*With AO*			*With RO*		
Estimator	*Mean*	*MSE*	*eff*	*Mean*	*MSE*	*eff*	*Mean*	*MSE*	*eff*
ML	−0.519	0.87	1.00	−0.122	16.07	1.00	−0.117	16.46	1.00
ILS	−0.518	0.88	0.98	−0.145	14.28	1.13	−0.115	16.73	0.98
IGM-a	−0.518	0.96	0.91	−0.491	1.55	10.38	−0.495	1.57	10.51
IGM-b	−0.518	0.96	0.91	−0.493	1.54	10.42	−0.492	1.58	10.39
RGM*		0.76/0.86			11.18/17.41				−/−

NOTES: Means and MSE are calculated on 500 replicate time series of length 100. Relative efficiencies (column eff) are ratios of the ML MSE to the estimator given in the line entrance. MSEs figures are multiplied by 100. See text for keys to symbols.

∗ Worst/best efficiencies obtained in Allende and Heiler's (1992) experiments with RGM estimators.

(IGM-a and IGM-b) happen to be robust to the contaminated situations considered with a slight advantage to IGM-a. Robustness is usually achieved at the cost of a loss of efficiency, although, in all the cases considered, this cost stays within reasonable limits. In Table 1 it can be seen that results under AO and RO are very similar so that only AO are considered in the remaining experiments. Finally, IGM performances are within the range of the one obtained by Allende and Heiler (1992) with RGM estimators.

Table 3 contains the results of the experiments performed with the ARMA(1,1) model. Results are, on the whole, similar to the ARMA(0,1) experiments. We note, however, a larger loss of efficiency from using ILS instead of ML estimation mainly for the autoregressive coefficient α. In nonreported experiments, much heavier loss was obtained when not using

Table 2. Monte Carlo Results for the Two ARMA(0,1) Cases

	Clean data			*With AO*		
Estimator	*Mean*	*MSE*	*eff*	*Mean*	*MSE*	*eff*
			$\beta = -0.8$			
ML	−0.823	0.57	1.00	−0.151	44.66	1.00
ILS	−0.804	0.61	0.93	−0.173	41.80	1.07
IGM-a	−0.788	0.70	0.81	−0.678	3.65	12.23
IGM-b	−0.789	0.69	0.82	−0.669	4.26	10.47
RGM*		0.79/0.98			11.58/24.44	
			$\beta = 0.5$			
ML	0.500	0.86	1.00	0.096	18.13	1.00
ILS	0.462	1.07	0.80	0.061	21.17	0.86
IGM-a	0.458	1.19	0.73	0.420	2.17	8.34
IGM-b	0.458	1.19	0.73	0.417	2.21	8.20

NOTES: Means and MSE are calculated on 500 replicate time series of length 100. Relative efficiencies (column eff) are ratios of the ML MSE to the estimator given in the line entrance. MSEs figures are multiplied by 100. See text for keys to symbols.

∗ Worst/best efficiencies obtained in Allende and Heiler's (1992) experiments with RGM estimators.

Table 3. Monte Carlo Results for the ARMA(1,1) Case With $\alpha = 0.8$ and $\beta = 0.5$

	Clean data			*With AO and τ^2 =9*			*With AO and τ^2 = 100*		
Estimator	*Mean*	*MSE*	*eff*	*Mean*	*MSE*	*eff*	*Mean*	*MSE*	*eff*
				Autoregressive parameter ($\alpha = 0.8$)					
ML	0.766	0.63	1.00	0.748	1.27	1.00	0.628	11.27	1.00
ILS	0.785	1.01	0.63	0.666	3.69	0.34	0.314	29.45	0.38
IGM-a	0.786	1.07	0.59	0.783	1.44	0.88	0.782	1.65	6.81
IGM-b	0.786	1.07	0.59	0.782	1.43	0.89	0.781	1.70	6.61
RGM*			0.67/1			3.07/5.80			–/–
				Moving average parameter ($\beta = 0.5$)					
ML	0.513	1.00	1.00	−0.017	31.93	1.00	−0.319	78.12	1.00
ILS	0.470	1.23	0.82	0.039	23.53	1.36	−0.027	29.84	2.62
IGM-a	0.466	1.33	0.75	0.344	5.14	6.21	0.418	3.17	24.67
IGM-b	0.467	1.34	0.75	0.341	5.22	6.12	0.418	3.13	24.98
RGM*			0.70/0.91			5.10/17.16			–/–

NOTES: Means and MSE are calculated on 500 replicate time series of length 100. Relative efficiencies (column eff) are ratios of the ML MSE to the estimator given in the line entrance. MSEs figures are multiplied by 100. See text for keys to symbols.

∗ Worst/best efficiencies obtained in Allende and Heiler's (1992) experiments with RGM estimators.

the instrumental variables described in Section 5.1. The efficiency loss for IGM remains here, however, within tolerable limits while the gain in robustness is important in most cases. An exception arises when estimating α in the AO case with $\tau^2 = 9$. We find it difficult to interpret the poor performance of IGM with respect to RGM in this particular case. A possible explanation might be the surprisingly robust estimate obtained with ML. This is not the case anymore when $\tau^2 = 100$. Note that in both of these cases ILS is much less robust than ML. Finally, we see that both IGM estimators are robust for α and β.

6. ILLUSTRATIVE EXAMPLES

We analyze two real time series, whose choice is motivated by the existence of previously published studies. This gives us ground for comparison of the proposed inferential method with approaches based on the detection of outliers. We do not tackle the model building issue in this article, and therefore simply use the models proposed in the previous available studies.

6.1 SAVING RATES SERIES

A time plot of the series on saving rates (saving as percent of income) in the United States from the first quarter of 1955 to the fourth quarter of 1979 is displayed in Figure 2. In Pankratz (1991, chap. 8) an outliers detection procedure inspired by Chang, Tiao, and Chen (1988) was applied to the series.

Chang et al.'s approach was based on an iterative procedure based on likelihood ratio tests to check the presence of IO, AO, or level shifts. These phenomena are then adjusted for (using intervention regressors) and the tests repeated. The procedure stops when tests are unable to detect any outliers. A similar procedure was proposed by Tsay (1986). This type of approach is most valuable if one is interested in the actual existence and nature of outliers in the time series. Such information may be used for historical analysis for instance. Thus,

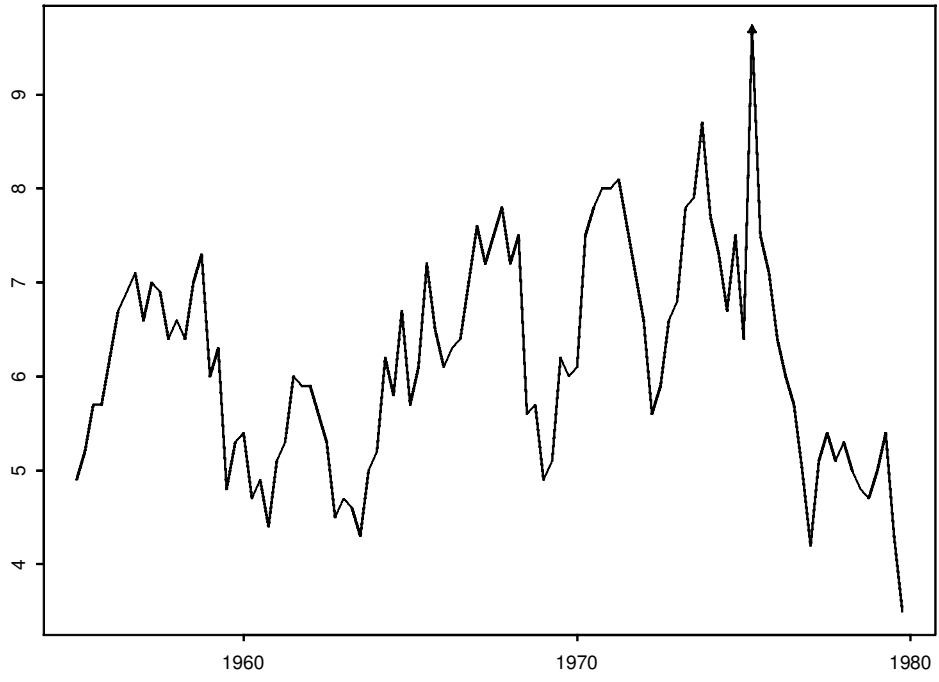

Figure 2. Time plot of the saving rates in the United States, from the first quarter of 1955 to the fourth quarter of 1979. A triangle highlights the 82nd (second quarter of 1975) observation.

Pankratz's (1991) analysis identified six time points with possible outlying phenomena, leading him to propose the following model for the saving rates series:

$$Y_t = \mu + w_{82}I_{1,t} + w_{99}I_{2,t} + w_{43}I_{3,t} + w_{89}I_{6,t} + W_t, \tag{6.1}$$

where $W_t = \alpha_1 W_{t-1} + X_t - \beta_2 X_{t-2}$, $X_t = w_{62}I_{4,t} + w_{55}I_{5,t} + Z_t$, and $I_{1,t}$, $I_{3,t}$, $I_{4,t}$, $I_{5,t}$, and $I_{6,t}$ are binary pulse variable equal to one at $t = 82$, 43, 62, 55, and 89, respectively, and zero elsewhere; $I_{2,t}$ is a binary step variable equal to zero for $t < 99$ and one for $t \geq 99$. The innovation process $\{Z_t\}$ is normally distributed with mean zero and variance σ^2. Model (6.1) is basically an ARMA(1,2) model with $\beta_1 = 0$, to which six interventions have been added to take into account detected outlying phenomena.

Our interest lies in the ARMA(1,2) actual parameters, α_1, β_2, μ, and σ^2. Table 4 reports the estimates obtained by Pankratz (1991) using model (6.1). However, Pankratz himself recognized that, for illustration purposes, he had used a hypersensitive outliers detection procedure. The probable consequence is that the estimate of the scale parameter σ^2 is overdeflated.

Looking at the saving rates series plot, it may be apparent that the 82nd observation (second quarter of 1975) is exceptionally large. Indeed, at that time the Congress passed a law granting a one-time tax rebate. This, according to some economic theories (see Pankratz 1991, chap. 7), may have caused the saving rates to rise. The 82nd observation may therefore be considered as an AO or RO. To adjust for this single event, Pankratz used the model (6.1) with $w_{43} = w_{55} = w_{62} = w_{89} = w_{99} = 0$, and obtained a larger estimate of σ^2 (see Table 4).

Table 4. ARMA(1,2) Parameters Fitted to the Saving Rates Data

	$\hat{\alpha}_1$		$\hat{\beta}_2$		$\hat{\mu}$	$\hat{\sigma}^2$
six interventions	0.80	(12.07)	0.38	(3.68)	6.16	0.23
one intervention	0.81	(11.76)	0.25	(2.23)	6.07	0.34
IGM	0.82	(7.71)	0.40	(2.39)	6.15	0.34
ML	0.74	(9.58)	0.34	(3.31)	6.11	0.44

NOTES: In parentheses are given absolute t values for the coefficients. The mean μ was clearly significant in all cases. The row entries indicate the estimation methods; from top to bottom, ML with six interventions, ML with a single intervention on the 82nd observation, the simulation-based estimation as described in the text, and ML without intervention. The t values for IGM are based on the variance estimator described in Remark 4.

An alternative inferential approach to the issue of fitting an ARMA(1,2) model to the saving rates series is to use a robustified procedure such as the one presented in this article. For this purpose, we used the simulation-based algorithm with normal pseudo-realizations in Step 2 and a GM estimator in Step 3 (S-Plus function `ar.gm` with `iterh=0`, `iterb=3` and `effgm=0.7`), $r = 3$ (AIC choice), and $s = 30$. The resulting estimates are reported in Table 4 together with estimates obtained with ML (S-Plus was used). Except for the scale parameter σ^2, we obtain similar results to Pankratz's estimates without the need of identifying the presence and nature of outliers. Maximum likelihood estimates of α_1 and β_2 are smaller in absolute value although not by much. The influence of the suspected outliers is most obvious with the ML estimate of σ^2, which seems overinflated. In this respect, the estimate of σ^2 obtained with our robust procedure is close to the one obtained with the single intervention model, while the coefficients (α_1 and β_2) estimates are similar to the ones obtained with model (6.1). A tentative interpretation is that model (6.1) has, as noted previously, too many interventions leading to an overdeflated scale estimate, while intervening too little, as with the single intervention model, lead to nonrobust coefficients estimators.

6.2 EXPORTS SERIES

We study here the monthly unadjusted series on exports from the USA to Latin-American republics between January 1966 and December 1983. This series has the interesting feature of containing patchy outliers. A plot of the logarithm of the series is given in Figure 3.

This dataset was used by Bruce and Martin (1989) to illustrate yet a different approach for the identification and treatment of outliers. Leave-k-out diagnostics were used to identify outliers while fitting an ARMA(0,2) model (normally distributed innovations were assumed) to the first differences of the logarithm of the series. Identified outliers were then deleted and treated as missing observations. The procedure is iterative in the sense that outliers are not all identified simultaneously, see Bruce and Martin (1989) for details.

For the export series their final analysis identified ten outliers at times 1/69–2/69 (for January 1969–February 1969), 9/71–11/71, 12/76–2/77, and 1/78–2/78. The first two groups of outliers (1/69–2/69 and 9/71–11/71) correspond to dock strikes and forestalling. The other groups of outliers have no known cause. The estimates obtained by ML when considering these time points as missing values are given in Table 5. In the same table, ML estimates

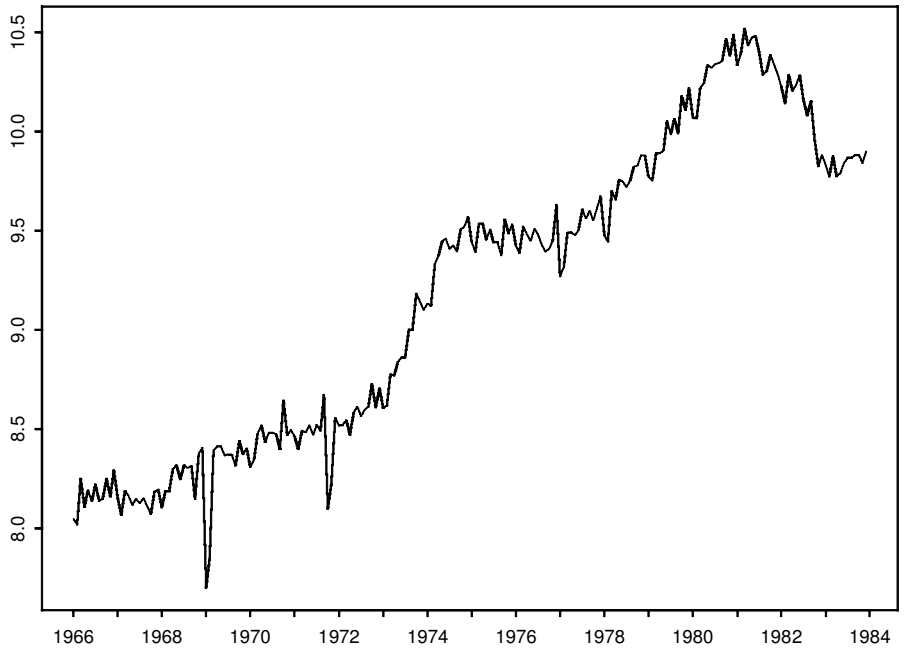

Figure 3. Time plot of the logarithm of monthly exports to Latin-America; starting January 1966 and ending December 1983.

based on all the observations are given, as well as estimates obtained with the herein proposed simulation-based algorithm. ML was performed with S-Plus and the simulation-based estimator used normal pseudo-realizations in Step 2 and a GM estimator in Step 3 (S-plus function `ar.gm` with `iterh=0`, `iterb=3` and `effgm=0.7`), $r = 2$ (AIC choice), and $s = 30$. We comment Table 5 by first noting that the simulation-based estimator (which avoids the cumbersome identification of outliers) gives almost identical results to the ML where Bruce and Martin's (1989) identified outliers are treated as missing observations. This is interesting for instance as a confirmation that no influential outlier has been omitted in Bruce and Martin's analysis. The results are also encouraging by showing a situation where the simulation-based estimator is robust to outliers occurring in groups. Note finally, that the ML estimator is strongly influenced by the outliers; a significant β_2 coefficient is obtained and the estimate of σ^2 is almost twice as large as the robust variants.

Table 5. ARMA(0,2) Parameters Fitted to the First Differences of the Logarithm of the Export Series

	$\hat{\beta}_1$		$\hat{\beta}_2$		$\hat{\mu}$	$\hat{\sigma}^2$
outliers deleted	0.43	(6.12)	0.08	(1.14)	0.008	0.0056
IGM	0.47	(6.87)	0.09	(0.97)	0.010	0.0060
ML	0.37	(5.11)	-0.16	(2.38)	0.008	0.0114

NOTES: In parentheses are given absolute t values for the coefficients. The mean μ was clearly nonsignificant in all cases. The row entries indicate the estimation methods; from top to bottom, ML with outliers treated as missing observations, the simulation-based estimation as described in the text, and ML without deletion of outliers. The t values for IGM are based on the variance estimator described in Remark 4.

7. DISCUSSION

A robust estimation method for the parameters of an ARMA model has been proposed in this article. A major advantage of the novel estimator over existing procedures, is that its asymptotic properties (distribution and influence function) can be derived. Moreover, a small sample study shows that the new estimator compares well with previously proposed methods. With two real examples, we have also been able to compare the proposed inferential method with two different approaches based on outliers detection.

The estimation algorithm as it is proposed in this article is essentially open in its structure; several of its components may be modified. For example, improvement of the robustness properties may be achieved by reconsidering the method with which the fit of the auxiliary model is performed. We have used the GM estimator in this article because it has been well studied in the literature. Possible alternatives to GM estimation could be the use of highly robust covariance estimators (Ma and Genton 2000) in conjunction with the Yule–Walker estimation equations, and the least median of squares estimator for regression parameters (Rousseeuw and Leroy 1987).

Moving average models are not specific to time series data but are of interest as models for any nonindependentlydistributed datasets. For instance, ARMA models are used to represent random fields in spatial statistics applications, see Cressie (1993). Thus, the proposed estimation algorithm could be adapted to such similar situations. More generally, the particular simulation-based estimation algorithm proposed and studied in this paper implicitly introduces a general robustifying framework for models for which no robust estimation criterion is available neither in closed form nor algorithmically; see also Genton and de Luna (2000) for general results concerning the influence function of robust simulation-based estimators.

Finally, we have not addressed the model selection stage although it is essential in the ARMA modeling of a time series. Model selection consists here in choosing the orders p and q in (2.1). For that purpose, automatic order selection via criteria such as AIC and BIC, which are basically penalized likelihoods (e.g., Brockwell and Davis 1991), is the common practice. By first evaluating the likelihood at the robust estimates of the parameters and then penalizing it, either as AIC or BIC, one should obtain a robust model selection criterion; see also Martin (1980) and Martin and Yohai (1986, rejoinder) for further discussion on this issue.

ACKNOWLEDGMENTS

The authors are grateful to Kurt Brännäs, Guido Kuersteiner, and Elvezio Ronchetti for their comments on an earlier version of the article. We thank also two anonymous referees, the associate editor, and the editor for valuable comments that helped to improve the article.

[Received August 1999. Revised May 2000.]

REFERENCES

Allende, H., and Heiler, S. (1992), "Recursive Generalized M Estimates for Autoregressive Moving-Average Models," *Journal of Time Series Analysis*, 13, 1–18.

Boente, G., Fraiman, R., and Yohai, V. J. (1987), "Qualitative Robustness for Stochastic Processes," *The Annals of Statistics*, 15, 1293–1312.

Brockwell, P. J., and Davis, R. A. (1991), *Time Series: Theory and Methods*, New York: Springer-Verlag.

Bruce, A. G., and Martin, R. D. (1989), "Leave-*k*-out Diagnostics for Time Series" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 51, 363–424.

Bustos, O. H. (1982), "General M-Estimates for Contaminated *p*th Order Autoregressive Processes: Consistency and Asymptotic Normality," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 59, 491–504.

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

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

Cressie, N. A. C. (1993), *Statistics for Spatial Data*, revised edition, New York: Wiley.

de Luna, X. (1998), "An Improvement of Akaike's FPE Criterion to Reduce its Variability," *Journal of Time Series Analysis*, 19, 457–472.

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

Diggle, P. J., and Gratton, R. J. (1984), "Monte Carlo Methods of Inference for Implicit Statistical Models" (with discussion), *Journal of the Royal Statistical Society*, Ser. B, 46, 193–227.

Genton, M. G. (1998), "Spatial Breakdown Point of Variogram Estimators," *Mathematical Geology*, 30, 853–871.

Genton, M. G., and de Luna, X. (2000), "Robust Simulation-based Estimation," *Statistics & Probability Letters*, 48, 253–259.

Gouriéroux, C., and Monfort, A. (1996), *Simulation-Based Econometric Methods*, Oxford: Oxford University Press.

Gouriéroux, C., Monfort, A., and Renault, A. E. (1993), "Indirect Inference," *Journal of Applied Econometrics*, 8, 85–118.

Hampel, F. R. (1974), "The Influence Curve and its Role in Robust Estimation," *Journal of the American Statistical Association*, 69, 383–393.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986), *Robust Statistics: The Approach based on Influence Functions*, New York: Wiley.

Huber, P. J. (1981), *Robust Statistics*, New York: Wiley.

——— (1984), "Finite Sample Breakdown Point of M- and P-Estimators," *The Annals of Statistics*, 12, 119–126.

Künsch, H. (1984), "Infinitesimal Robustness for Autoregressive Processes," *The Annals of Statistics*, 12, 843–863.

Ma, Y., and Genton, M. G. (2000), "Highly Robust Estimation of the Autocovariance Function," *Journal of Time Series Analysis*, 21, 663–684.

Mallows, C. L. (1976), "On some Topics in Robustness," *Bell Labs Tech. Memo*, Murray Hill, NJ.

Martin, R. D. (1980), "Robust Estimation of Autoregressive Models," in *Directions in Time Series*, eds. D. R. Brillinger and G. C. Tiao, Hayward, CA: Institute of Mathematical Statistics, pp. 228–254.

Martin, R. D., and Jong, J. (1977), "Asymptotic Properties of Robust Generalized M-Estimates for the First-Order Autoregressive Parameter," Bell Labs Techncial Memo, Murray Hill, NJ.

Martin, R. D., and Yohai, V. J. (1985), "Robustness in Time Series and Estimating ARMA Models," in *Handbook of Statistics 5*, eds. E. J. Hannan, P. R. Krishnaiah, M. M. Rao, pp. 119–155.

——— (1986), "Influence Functionals for Time Series" (with discussion), *The Annals of Statistics*, 14, 781–855.

Pankratz, A. (1991), *Forecasting with Dynamic Regression Models*, New York: Wiley.

Rousseeuw, P. J., and Leroy, A. M. (1987), *Robust Regression and Outliers Detection*, New York: Wiley.

Rubinstein, R. Y. (1981), *Simulation and the Monte Carlo Method*, New York: Wiley.

Tsay, R. S. (1986), "Time Series Model Specification in the Presence of Outliers," *Journal of the American Statistical Association*, 81, 132–141.