

Recursive Generalized M-Estimators of System Parameters

Carlo GRILLENZONI

IUAV: University Institute of Architecture of Venice
30135 Venice
Italy
(carlog@brezza.iuav.unive.it)

This article develops recursive versions of generalized M-estimators for the parameters of dynamic systems. The starting point is provided by the robust algorithms for autoregressive moving average models proposed in the past decade. Using parallel calculation of "cleaned" input and output, bounded influence regression is applied recursively. The resulting estimators have the same structure as the weighted least squares but use filtered regressors. Monte Carlo experiments are carried out to check the robustness of the proposed estimators and to compare their performance with that of other methods. Numerical results are very encouraging.

KEY WORDS: Additive outliers; Recursive least squares; Robust regression; Transfer function models; Winsorized processes.

Anomalous observations (outliers) may occur in time series data for several reasons, such as measurement errors, structural breaks, interventions, and so forth. Their presence can seriously affect least squares (LS) and maximum likelihood estimates of the parameters of dynamic models, making their use in forecasting and control unreliable.

There are two major approaches by which outliers in time series can be handled. One is by *robust* estimators (e.g., Martin and Yohai 1985) and the other is by *detection* techniques (e.g., Chang, Tiao, and Chen 1988; Abraham and Chuang 1989). Robust methods keep fixed the structure of the models and aim to provide estimates that have small bias in the presence of outliers and are nearly optimal in the absence of them. On the other hand, detection techniques use tests and diagnostics for identifying the position and the nature of outliers and then accommodate their effects by introducing dummy variables into the models.

In certain contexts, like adaptive control of engineering processes, there is need for *recursive* (on-line) procedures for outliers. By this I mean methods that update the information as more data become available, allowing for fast calculation. Even though a detection method based on time series recursions was recently developed by Ljung (1993) the on-line environment is more suitable for robust estimators. Indeed, recursive algorithms of the M-type (in the sense of Huber 1981) have been developed in the estimation of system parameters (Ljung and Söderström 1983), location coefficients (Cameron and Turner 1993), and state-space models (Schick and Mitter 1994).

It is well known (Martin and Yohai 1985, p. 132) that M-estimates are robust against innovation outliers (IO's) but are not robust against additive outliers (AO's). The first occur in the disturbances and in the inputs and affect the output through the dynamic of the system but the second are typically associated with measurement errors of both input and output series. To cope with this problem, the solution of *generalized* M-estimates (GM), a form of bounded influence regression, was developed in time series analysis by Denby and Martin (1979) and Martin (1980).

This article aims to derive recursive GM-estimators for the parameters of the transfer function models of Box and Jenkins (1976). The starting point is provided by the *residual autocovariance* algorithm designed by Bustos and Yohai (1986) for univariate models. This belongs to the class of GM-estimators, and its attractive feature is the application of standard time series procedures to "cleaned" observations. After proper modifications, it is therefore suitable for the recursive implementation.

With respect to the classical iterative algorithms used in robust estimation, the recursive implementation makes calculation faster because it avoids direct optimization. More generally, it enables the dynamic effects of outliers to be monitored and changes in the regression parameters to be tracked (see Grillenzoni 1994). This is achieved, however, at the cost of increasing the error variance for finite sample size.

The article is organized as follows: Section 1 provides background material for nonlinear regression; Section 2 is concerned with basic aspects of recursive M-estimation; Section 3 deals with the central topic of the article; finally, Section 4 applies recursive GM-estimators to simulated data. Throughout, applications to real data are presented.

1. ITERATIVE LS ESTIMATES

Transfer function models developed by Box and Jenkins (1976, part III) connect an output process $\{y_t\}$ to an input process (possibly a control input) $\{x_t\}$ in the following way:

$$y_t = \frac{(\omega_0 + \omega_1 B + \dots + \omega_s B^s)}{(1 - \delta_1 B - \dots - \delta_r B^r)} x_{t-b} + \frac{(1 + \theta_1 B + \dots + \theta_q B^q)}{(1 - \phi_1 B - \dots - \phi_p B^p)} a_t, \quad a_t \sim \text{IN}(0, \sigma^2), \quad (1a)$$

© 1997 American Statistical Association
and the American Society for Quality Control
TECHNOMETRICS, MAY 1997, VOL. 39, NO. 2

and

$$(1 - \phi_{x1}B - \dots - \phi_{xp_x}B^{p_x})x_t = (1 + \theta_{x1}B + \dots + \theta_{xq_x}B^{q_x})e_t, \quad e_t \sim \text{IN}(0, \sigma_e^2), \quad (1b)$$

where $E(x_t a_s) = 0$ for all t, s and $\{x_t\}$ has an autoregressive moving average (ARMA) representation. The preceding notations mean that $\{a_t, e_t\}$ are independent normal (IN) disturbances, $\{\omega_i, \delta_j, \theta_i, \phi_j; \theta_{xi}, \phi_{xj}\}$ are constant parameters, and B is the backward-shift operator ($B^b x_t = x_{t-b}$).

Using polynomial operators of the type $\alpha(B) = (1 + \alpha_1 B + \dots + \alpha_k B^k)$, model (1b) may be rewritten as $\phi_x(B)x_t = \theta_x(B)e_t$ and (1a) may be decomposed as $y_t = m_t + n_t$, where $m_t = [\omega(B)/\delta(B)]x_{t-b}$ and $n_t = [\theta(B)/\phi(B)]a_t$ are independent subsystems. Conditions of stationarity and invertibility then require that all polynomials have roots outside the unit circle (i.e., stable roots), except $\omega(B)$, which only needs to have bounded coefficients.

For the developments of the next sections it is useful to represent model (1a) in "regression" form. This follows by noting that $m_t = (\delta_1 m_{t-1} + \dots + \delta_r m_{t-r}) + \omega(B)x_{t-b}$ and $n_t = (\phi_1 n_{t-1} + \dots + \phi_p n_{t-p}) + \theta(B)a_t$; hence, using $y_t = m_t + n_t$, one may obtain

$$y_t = \left(\sum_{i=1}^r \delta_i m_{t-i} + \sum_{j=0}^s \omega_j x_{t-b-j} + \sum_{i=1}^p \phi_i n_{t-i} + \sum_{j=1}^q \theta_j a_{t-j} \right) + a_t \stackrel{\text{def}}{=} \beta' z_t + a_t, \quad (2)$$

where $\beta' = [\dots \delta_i \dots \omega_j \dots \phi_i \dots \theta_j \dots]$ is the vector of parameters and $z_t' = [\dots m_{t-i} \dots x_{t-b-j} \dots n_{t-i} \dots a_{t-j} \dots]$ is the vector of "regressors." For given β , the entries of z_t may be generated from the processes $\{y_t, x_t\}$, as discussed by Box and Jenkins (1976, p. 389); therefore, the "residual of regression" associated with model (2) can be written as $a_t(\beta) = [y_t - \beta' z_t(\beta)]$ where $z_t(\cdot)$ itself is a function of the parameters.

Given a sample $\{y_t, x_t; t = 1, 2, \dots, T\}$, the LS estimator of β is the value that minimizes the functional $Q_T(\beta) = \sum_{t=1}^T a_t^2(\beta)$. Following a Gauss-Newton strategy, the residual's gradient must be defined; from Grillenzoni (1991), this is given by

$$\xi_t(\beta) = -\frac{\partial a_t(\beta)}{\partial \beta} = G(B)z_t(\beta)$$

$$G(B) = \text{diag} \left[\frac{\phi(B)}{\delta(B)\theta(B)} \mathbf{I}_{(r+s+1)}; \frac{1}{\theta(B)} \mathbf{I}_{(p+q)} \right], \quad (3)$$

where $\mathbf{I}_{(r+s+1)}$ and $\mathbf{I}_{(p+q)}$ denote identity matrices of sizes $(r + s + 1)$ and $(p + q)$. From (3) it is clear that the gradient $\{\xi_t\}$ is obtained by filtering the regressors $\{z_t\}$ with a stable filter $G(B)$ that depends on the parameters β . Moreover, under Gaussianity, it is easy to prove that

the expected value of the Hessian of $Q_T(\beta)$ is given by $E(\xi_t \xi_t') \cdot T$.

The Gauss-Newton algorithm in the $(k + 1)$ th iteration then becomes

$$\text{LS} \quad \hat{\beta}_T(k+1) = \hat{\beta}_T(k) + \left[\sum_{t=b+1}^T \hat{\xi}_t(k) \hat{\xi}_t(k)' \right]^{-1} \sum_{t=b+1}^T \hat{\xi}_t(k) \hat{a}_t(k) \quad (4a)$$

$$\hat{\xi}_t(k) = \hat{G}_T(k, B) \hat{z}_t(k)$$

$$\hat{a}_t(k) = [y_t - \hat{\beta}_T(k)' \hat{z}_t(k)], \quad (4b)$$

where $\hat{a}_t(k)$ are residuals, $\hat{z}_t(k)$ are the regressors generated by $\hat{\beta}_T(k)$, and $\hat{G}_T(k, B)$ is the filter (3) evaluated at the same point (for details, see Grillenzoni 1991).

Under the assumptions of stationarity and invertibility of (1a), the estimator (4) enjoys optimal asymptotic properties; in particular, following Pierce (1972), I have

$$\sqrt{T}[\hat{\beta}_T(k) - \beta] \xrightarrow{L} N[0, E(\xi_t \xi_t')^{-1} \sigma^2]$$

as $(k, T) \rightarrow \infty$. (5)

Note that, under the same assumptions, the processes $\{z_t, \xi_t\}$ are stationary and the dispersion in (5) does not depend on t . In particular, in the univariate case it turns out to be a complicated function of β (see Martin and Yohai 1985, p. 127).

Bustos and Yohai (1986) argued that when outliers are of innovation type—that is, when the density of $\{a_t\}$ has heavy tails (and its functional form is known a priori)—the LS estimator becomes *inefficient* compared with that of maximum likelihood. On the other hand, a considerable bias in both estimators arises when the processes $\{y_t, x_t\}$ are affected by AO's (see Martin 1980). A contamination scheme by AO for model (1) is given by

$$y_t^o = y_t + u_t v_t, \quad u_t \sim \text{IB}(1, \pi_u),$$

$$v_t \sim \text{IN}(0, \sigma_v^2), \quad (6a)$$

and

$$x_t^o = x_t + z_t w_t, \quad z_t \sim \text{IB}(1, \pi_z),$$

$$w_t \sim \text{IN}(0, \sigma_w^2), \quad (6b)$$

where $\{u_t, z_t\}$ are Bernoulli (0-1) processes and $\{v_t, w_t\}$ have variances larger than those of $\{y_t, x_t\}$, all the processes being independent. In practice, with probability $1 - \pi_u = P(u_t = 0)$ the output y_t is itself observed, but something different is observed with probability $\pi_u = P(u_t = 1)$. Scheme (6) may also be used for defining the contamination by IO in the same system; this follows by replacing in (1) the processes $\{a_t, e_t\}$ with $a_t^o = a_t + u_t v_t$ and $e_t^o = e_t + z_t w_t$.

The previous outlier classification is typical of univariate time series. In the multivariate context, however, the separation between additive and innovation outliers is not so precise. For example, if I have an AO in the input series that is due to a real change, then the effect on the output would be similar to that of an IO. For the sake of simplicity,

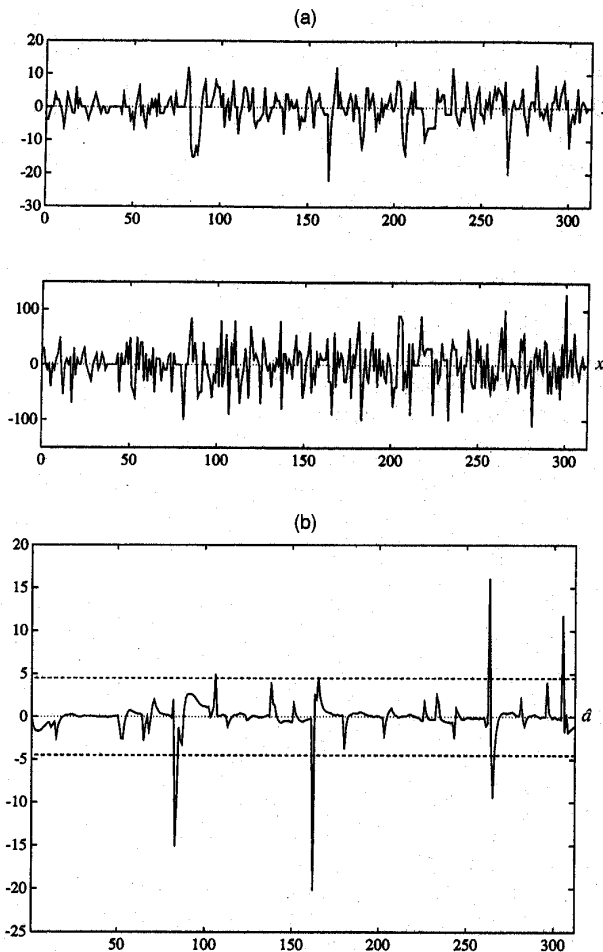


Figure 1. Pilot Scheme Data: (a) Series y_t, x_t ; (b) Residuals \hat{a}_t of Model (7).

in the following I only consider AO's associated with measurement errors, hence affecting separately the two series $\{x_t, y_t\}$.

Application 1. To introduce the problem of robust estimation, I begin an application to the dataset of Box and Jenkins (1976) named "Pilot Scheme." The case study involves fitting a transfer function model from the operating data of a control experiment in which the output is a polymer viscosity and the input is the gas rate. Box and Jenkins (1976, p. 451) argued that this is possible by taking as series $\{y_t\}$ the control errors (i.e., the deviations from target of the actual viscosity) and as series $\{x_t\}$, the changes in the gas rate. These series are shown in Figure 1(a).

If one repeats the exercise of Box and Jenkins (1976, pp. 451-459), including in the system the contemporaneous term $\omega_0 x_t$, one obtains

$$y_t = \begin{pmatrix} 1 - .517B \\ (12.1) \end{pmatrix}^{-1} \begin{pmatrix} -.102 + .103B \\ (26.7) \end{pmatrix} x_t + \begin{pmatrix} 1 - .872B \\ (37.2) \end{pmatrix} \hat{a}_t, \quad \hat{\sigma} = 2.25, \quad (7)$$

where the values in parentheses are t statistics. The preceding model has smaller residual variance than that of Box and Jenkins. A plot of the residuals, however, reveals the presence of some large outliers [see Fig. 1(b)]. This makes the estimates in (7) unreliable.

In standard regression, a simple method of solving the problem of outliers consists of discarding observations corresponding to big residuals. To preserve the autocorrelation structure in time series, a related approach replaces the outlying data with their conditional means, given the remaining observations (see Ljung 1993). This is particularly effective for AO's because IO's theoretically affect all subsequent observations. If the noise component has an MA(1) structure, however, as in (7), the effects of an a_t^o only concern the pair y_t, y_{t+1} .

I tentatively replaced the data at $t = 83, 85; 162, 163; 263, 264; 305, 306$ with the conditional means and reestimated model (7). As a result the residual plot was satisfactory, but the value of the MA parameter became negligible— $\hat{\theta}_1 = -.094, \hat{\sigma} = 1.045$; the other estimates remained nearly unchanged— $\hat{\delta}_1 = .506, \hat{\omega}_0 = -.101, \hat{\omega}_1 = .102$. These facts suggest that outliers in the Pilot Scheme data may be both of innovation and additive type; in particular, at $t = 83, 85$ one might have two AO's but the remaining might be IO's.

In the following I shall use the preceding results as benchmarks for the evaluation of the performance of alternative robust estimators. In general, changes in the parameter estimates could be used as diagnostic tools for outlier detection; however, test procedures, as given by Chang et al. (1988) or Abraham and Chuang (1989), are more powerful for this purpose.

Application 2. The second application focuses on another industrial dataset published by Box and Jenkins (1976, p. 532) and named "Gas Furnace." The case study concerns a gas furnace in which air and methane were combined to form a mixture of gases containing carbon dioxide (CO_2). In this system, the air feed was kept constant, but the methane feed rate (input) was varied and the resulting CO_2 concentration (output) measured. These series were investigated for the presence of outliers by Peña (1991), who found influential observations around $t = 270$. On the other hand, Grillenzoni (1994) showed that the parameters of the model identified by Box and Jenkins (1976, p. 381) become time-varying in the second part of the sample, starting from $t = 167$.

In this application I am interested in studying the effects of AO's; therefore, I have considered only the first 166 observations. For this subsample, the model specified by Box and Jenkins simplifies as

$$y_t = \begin{pmatrix} 1 - .694B \\ (118.2) \end{pmatrix}^{-1} \begin{pmatrix} -1.032B^3 \\ (60.5) \end{pmatrix} x_t + \begin{pmatrix} 1 - .782B \\ (19.4) \end{pmatrix}^{-1} \begin{pmatrix} 1 + .347B \\ (5.2) \end{pmatrix} \hat{a}_t, \quad \hat{\sigma} = .141, \quad (8)$$

and $\{x_t\}$ becomes an AR(2) process $(1 - 1.69B + .765B^2)x_t = e_t, \hat{\sigma}_e = .219$.

Introducing the outliers $y_{80}^0 = -16, x_{110}^0 = +6$, which are nearly twice the maximum values of the series, the LS estimates become $\hat{\delta}_1 = .750, \hat{\omega}_0 = -.853, \hat{\phi}_1 = .737$, and $\hat{\theta}_1 = -.516$. As in application 1, one sees that greatest effects are on the MA parameter. In Section 2, I evaluate the performance of several robust estimators, keeping as a benchmark model (8).

2. RECURSIVE M-ESTIMATES

Discussion of M-estimates is a necessary introduction to robust regression. The approach of Huber (1981) was to modify the loss function of the LS estimator in such a way as to downweight anomalous observations. Letting $P_T(\beta) = \sum_{t=1}^T \rho(a_t(\beta))$, with $\rho(\cdot)$ a differentiable symmetric and convex function, M-estimates arise by numerical minimization of $P_T(\cdot)$. Alternatively, one may solve iteratively the system of normal equations

$$\partial P_T(\beta) / \partial \beta = \sum_{t=1}^T \psi(a_t(\beta)) \xi_t(\beta) = 0, \tag{9}$$

where $\psi(x) = \partial \rho(x) / \partial x$ is a bounded and odd function.

In Huber's approach, this function has the form $\psi_H(x) = \text{sign}(x) \cdot \min\{|x|, c\}$, where c is a design constant that must be selected according to the rate of outlier contamination. To avoid dependence on scale factors, the term $a_t(\beta)$ may be replaced by (a_t/σ) ; in this case, the parameter σ can be estimated by introducing in (9) the further equation $\sum_{t=1}^T \chi(a_t/\sigma) = 0$, where $\chi(\cdot)$ is an even function that must be related to $\psi(\cdot)$.

Computation of robust estimates is based on iterative algorithms that go back to Beaton and Tukey (1974) for which recursive versions do not exist. Development of on-line methods is needed for increasing the speed of calculation but also for monitoring the dynamic effects of outliers. This may be useful for identifying their exact nature.

The derivation of the recursive M-estimator for (1) requires that Huber's solution be formulated in terms of weighted least squares; namely, $Q_T^*(\beta) = \sum_{t=1}^T [w_t \cdot a_t(\beta)]^2$, where $0 < w_t < 1$. This is enabled by the algebraic relationships between the two methods; in particular, choosing weights of the form $w(x) = \psi(x)/x$ with $x = a_t/\sigma$, one may easily obtain system (9). With the new formulation, however, it is possible to derive an explicit expression for the iterative algorithm that minimizes $P_T(\beta)$. This is simply given by (4) with the gradients $\hat{\xi}_t(k)$ multiplied by w_t^2 .

Once the M-estimates have been expressed in this general form, the transformations of Ljung and Söderström (1983, chap. 2) directly yield the recursive M-algorithm

$$\begin{aligned} \text{RM} \quad \hat{\beta}(t) &= \hat{\beta}(t-1) + \mathbf{R}(t)^{-1} w_t^2 \cdot \hat{\xi}(t) \hat{a}(t), \\ \hat{a}(t) &= y_t - \hat{z}(t)' \hat{\beta}(t-1), \end{aligned} \tag{10a}$$

$$\begin{aligned} \mathbf{R}(t) &= \mathbf{R}(t-1) + w_t^2 \cdot \hat{\xi}(t) \hat{\xi}(t)', \\ \hat{\xi}(t) &= \hat{\mathbf{G}}(t-1, B) \hat{z}(t). \end{aligned} \tag{10b}$$

In (10), $\hat{a}(t)$ are prediction errors, $\hat{z}(t)' = [\hat{m}(t-1) \dots \hat{a}(t-q)]$ is the vector of regressors, $\mathbf{R}(t)$ is the Hessian, and $\hat{\mathbf{G}}(t-1, B)$ is the filter (3) evaluated at $\hat{\beta}(t-1)$. Computational details for the vectors $\hat{z}(t)$ and $\hat{\xi}(t)$ were given by Grillenzoni (1991).

The choice of the weights may be related to the psi-function by recalling that $w(x) = \psi(x)/x$. In particular, Huber's solution, $\psi_H(x)$ and $x = a_t/\sigma$, yields

$$\hat{w}_c(t) = \begin{cases} 1 & \text{if } |\hat{a}(t)| \leq c \hat{\sigma}(t-1) \\ c \hat{\sigma}(t-1) |\hat{a}(t)|^{-1} & \text{if } |\hat{a}(t)| > c \hat{\sigma}(t-1), \end{cases} \tag{11}$$

where $\hat{\sigma}(t)$ is a recursive estimator of the scale and (c) is the design constant.

A robust estimator for σ may be based on Winsorized prediction errors, defined as $\hat{a}^*(t) = \psi_H[\hat{a}(t)/\hat{\sigma}(t-1)]\hat{\sigma}(t-1)$ [see Appendix A(1)]. Given the relationship $w(x) = \psi(x)/x$, it is easy to check the correspondence $\hat{a}^*(t) = \hat{w}_c(t)\hat{a}(t)$; with this, the recursive version of the robust estimator $\hat{\sigma}_T^2 = T^{-1} \hat{Q}_T^*$ becomes

$$\hat{\sigma}^2(t) = \frac{1}{t} [(t-1)\hat{\sigma}^2(t-1)] + \frac{1}{t} [\hat{w}_c(t)\hat{a}(t)]^2. \tag{12}$$

This solution will be referred to as MSE and completes the recursive M-algorithm (10)–(11).

The preceding formulation does not hold for other kinds of psi-functions, such as the *bisquare* one proposed by Beaton and Tukey (1974) [see Appendix A(2)]. The reason is that Winsorization is a concept closely related to $\psi_H(\cdot)$. The formal justification of (12) stems from the fact that in Huber's approach the chi-function has the form $\chi_H(x) = \psi_H^2(x) - \alpha$, where α has to allow consistency in the Gaussian case, namely, $\alpha = E[\psi_H^2(x)]$ for $x \sim N(0, 1)$. Thus, setting $x = a_t/\sigma$ and $\alpha = 1$, from the equation $\sum_{i=1}^t \chi(a_i/\sigma) = 0$, one may obtain (12). Finally, it is worth recalling that in the absence of outliers with $w_t = 1$, the estimator $\hat{\sigma}(t)$ is asymptotically unbiased [see Appendix A(3)].

A solution for $\hat{\sigma}_T$ that is popular in robust applications is the median absolute deviation (MAD) of residuals, defined as $\text{MAD}(\hat{a}_T) = \text{med}|\hat{a}_t - \text{med}(\hat{a}_t)|$, where $b+1 \leq t \leq T$ and med is the median. In sequential form we can define

$$\hat{\sigma}_t = \text{MAD}(\hat{a}_t) / .6745, \tag{13}$$

where the constant .6745 is chosen so that for Gaussian errors the preceding is a consistent estimator of σ (see Hogg 1977). With respect to (12), the statistic (13) slows down the computation speed in (10) but does not require the initial value $\hat{\sigma}(0)$. A way of reducing the amount of extra computation in (13) consists of using running medians of 3 as discussed by Gebski and McNeil (1984).

3. RECURSIVE GM-ESTIMATES

Typically, the need for GM-estimators arises when the observed series $\{x_t, y_t\}$ are affected by AO's. In Model (1) the contamination extends to the vectors $\{z_t, \xi_t\}$, and the related M-estimates become nonrobust. The classical solution is based on System (9) and consists of enlarging the domain of the function $\psi(\cdot)$ so as to include regression variables (see Martin and Yohai 1985, p. 134). Hence, denoting

the enlarged psi-function by $\eta(\cdot, \cdot)$, the modified system (9) becomes

$$\sum_{t=1}^T \eta[a_t(\beta)/\sigma, \xi_t(\beta)] \xi_t(\beta) = 0.$$

A suitable choice for eta is the Mallows function $\eta_M(x, y) = \psi_1(x)\psi_2(y)$, where $\psi_1(\cdot), \psi_2(\cdot)$ are univariate and multivariate psi-functions, respectively. The combination with Huber's $\psi_H(\cdot)$ makes $\eta_M(\cdot, \cdot)$ odd in each variable.

The derivation of the recursive GM-estimator for model (1) requires a simplification of the classical approach. Looking at the previous set of equations, one realizes that another solution may be obtained from the system $\sum_{t=1}^T \eta[a_t(\beta)/\sigma, \xi_t(\beta)] = 0$, where $\eta(\cdot, \cdot)$ is a suitable vector function. This approach provides, in fact, a natural extension of the M-estimation and has some connection with the approach of Bustos and Yohai (1986). In particular, using a vector of Mallows functions $\eta(\cdot, \cdot)' = [\eta_M(\cdot, \cdot) \dots \eta_M(\cdot, \cdot)]$, for each component $i = 1, 2, \dots, (r + s + 1 + p + q)$, one may obtain

$$\begin{aligned} \sum_{t=1}^T \eta_M \left(\frac{a_t}{\sigma}, \frac{\xi_{it}}{\sigma_i} \right) \sigma \sigma_i &= \sum_{t=1}^T \psi_H \left(\frac{a_t}{\sigma} \right) \sigma \psi_H \left(\frac{\xi_{it}}{\sigma_i} \right) \sigma_i \\ &= \sum_{t=1}^T \xi_{it}^* a_t^* = 0, \end{aligned} \quad (14)$$

where σ_i are standard deviations of ξ_{it} and $\{\xi_{it}^*, a_t^*\}$ are Winsorized processes. Now, system (14) provides a set of normal equations that are similar to $\partial Q_T / \partial \beta = \sum_t \xi_t a_t = 0$; the corresponding iterative algorithm is therefore given by (4) with $\xi_t(k), \hat{a}_t(k)$ replaced by $\xi_t^*(k), \hat{a}_t^*(k)$.

Given the relationship $\xi_t^* = G(B)z_t^*$, the computation of the implied GM-estimator requires the vector of cleaned regressors $z_t^* = [\dots n_{t-i}^* \dots x_{t-b-j}^* \dots n_{t-i}^* \dots a_{t-j}^* \dots]'$. A way of obtaining $\{z_t^*\}$, the cleaned input sequence, comes from the iterative procedure proposed by Bustos and Yohai (1986). The method computes residual autocovariance (RA) estimates for ARMA parameters; with respect to model (1b), it can be summarized as follows:

Step 1. Given initial estimates $[\hat{\phi}_{xi}, \hat{\theta}_{xj}, \hat{\sigma}_e]$, calculate the residuals $\hat{e}_t = \hat{\phi}_x(B)^{-1} \hat{\theta}_x(B) x_t$ for each $t = (p_x + 1) \dots T$ and clean them as $\hat{e}_t^* = \psi(\hat{e}_t / \hat{\sigma}_e) \hat{\sigma}_e$.

Step 2. Generate the pseudoserries $\hat{x}_t^* = \hat{\theta}_x(B)^{-1} \hat{\phi}_x(B) \hat{e}_t^*$ and reestimate the parameters $[\phi_{xi}, \theta_{xj}, \sigma_e]$ by applying the Gauss-Newton algorithm to $\{\hat{x}_t^*\}$.

The steps are iterated until convergence of estimates is achieved. Bustos and Yohai (1986) argued that the method is consistent if $\{x_t\}$ has no outliers. It is robust in the presence of AO's only for autoregressive models, however, because in Step 1 the effects of an outlier $x_t^?$ spread out on the whole sequence $\hat{e}_{t+h}, h > 0$. As a solution to this drawback, the authors proposed truncated RA estimates, which arise by approximating the function $\theta_x^{-1}(B)\phi_x(B)$ with a finite polynomial. Such an approach is not optimal and cannot be applied if $\theta_x(B)$ has roots near the unit circle.

By contrast, an "exact" solution can be obtained from parallel calculation of the series $\{e_t^*, x_t^*\}$. Basically, this means that, given $e_t^* = \psi(e_t/\sigma_e)\sigma_e$ at time (t) , the quantity x_t^* is immediately computed as a linear prediction: $x_t^* = (\sum_i \phi_{xi} x_{t-i}^* + \sum_j \theta_{xj} e_{t-j}^*) + e_t^*$. Next, at time $(t + 1)$, one calculates $e_{t+1} = x_{t+1} - (\sum_i \phi_{xi} x_{t+1-i}^* + \sum_j \theta_{xj} e_{t+1-j}^*)$, which only depends on the outlier $x_{t+1}^?$. The advantage in terms of uniform robustization is apparent.

In Appendix B, these ideas are extended to the GM-estimation of the parameters of transfer-function models. An iterative algorithm of Huber-Mallows type is described in detail.

Parallel calculation of cleaned input-output series is a necessity in the derivation of recursive GM-estimators for dynamic systems. To simplify the exposition, I assume that $\{y_t, x_t\}$ have the representation $\phi(B)y_t = \omega(B)x_{t-b} + \theta(B)a_t$, or I convert model (1a) in this structure by multiplying by $\tilde{\phi}(B) = \phi(B)\delta(B)$. Hence, combining the preceding discussion on the calculation of $\{e_t^*, x_t^*\}$ with the recursive M-estimator (10), one may obtain

$$\text{RGM } \hat{a}(t) = y_t - \hat{z}^*(t)' \hat{\beta}(t-1), \quad (15a)$$

$$\mathbf{R}(t) = \mathbf{R}(t-1) + \hat{w}_c^2(t) \hat{\xi}^*(t) \hat{\xi}^{*'}(t)', \quad (15b)$$

$$\hat{\beta}(t) = \hat{\beta}(t-1) + \mathbf{R}^{-1}(t) \hat{w}_c^2(t) \hat{\xi}^*(t) \hat{a}(t), \quad (15c)$$

$$\hat{a}(t) = y_t - \hat{\beta}(t)' \hat{z}^*(t), \quad (15d)$$

$$\begin{aligned} \hat{a}^*(t) &= \psi_H[\hat{a}(t)/\hat{\sigma}(t-1)] \hat{\sigma}(t-1) \\ &= \hat{w}_c(t) \hat{a}(t), \end{aligned} \quad (15e)$$

$$\hat{y}^*(t) = \hat{\beta}(t)' \hat{z}^*(t) + \hat{a}^*(t), \quad (15f)$$

$$\begin{aligned} \hat{z}^*(t+1) &= [\hat{y}^*(t) \dots \hat{x}^*(t-b-i) \\ &\dots \hat{a}^*(t-q+1)]', \end{aligned} \quad (15g)$$

$$\hat{\xi}^*(t+1) = \hat{z}^*(t+1) - \sum_{j=1}^q \hat{\theta}_j(t) \hat{\xi}^*(t+1-j), \quad (15h)$$

where $\hat{w}_c(t)$ and $\hat{\sigma}(t)$ are computed as in (11) and (12), respectively. The extension to transfer function models can be obtained from the algorithm in Appendix B; only one modification concerns the calculation of the vectors $\hat{z}^*(t)$ and $\hat{\xi}^*(t)$ in Equations (15g) and (15h).

Some technical explanations are necessary to motivate (15):

1. The recalculation of the prediction error in Equation (15d) is made for improving the convergence of the estimator $\hat{\beta}(t)$ (see Ljung and Söderström 1983, p. 316).

2. The introduction of the factor $\hat{w}_c^2(t)$ in the equations of $\hat{\beta}(t)$ and $\mathbf{R}(t)$ is motivated by the goal of encompassing the M-estimator and minimizing $Q_t^* = \sum_{i=1}^t (a_i^*)^2$.

3. The structure of the gradient (15h) arises from the fact that, when $\phi(B) = \delta(B)$, one has $G(B) = \theta^{-1}(B) \mathbf{I}_{(p+s+q)}$ and consequently $\theta(B)\xi_t = z_t$.

Table 1. Pilot Scheme Dataset: Parameter Estimates of Model (7)

Method	Scale	δ_1	ω_0	ω_1	θ_1	σ
LS*		.506	-.101	.102	-.094	1.045
LS		.517	-.102	.103	-.872	2.251
M	MSE	.500	-.100	.100	+.282	.297
GM	"	.512	-.102	.103	+.123	.798
RLS ($t = T$)	MSE	.521	-.098	.101	-.763	2.35
RM ($t = T$)	"	.525	-.099	.100	-.314	.671
RGM ($t = T$)	"	.524	-.102	.103	-.049	.653

NOTE: LS* = parameters estimated on data with outliers replaced by conditional means.

4. Computation of $\hat{x}^*(t - b - i)$ in (15g) is accomplished by an algorithm that has the same structure as (15) and is simultaneously applied to the model for $\{x_t\}$.

Application 1. Table 1 summarizes the results of applying both the iterative and the recursive estimators to model (7) of the pilot scheme data. From Appendix B, iterative M-estimates were obtained with a procedure that alternates the minimization of $Q_T^* = \sum_{t=1}^T (a_t^*)^2$ and the computation of $\hat{\sigma}^2 = T^{-1} \hat{Q}_T^*$. GM-estimates differ inasmuch as they use cleaned regressors in generating \hat{a}_t . Because the input is

a white noise, the related cleaned series was obtained as $\hat{x}_t^* = \psi_H(x_t / \hat{\sigma}_x) \hat{\sigma}_x$. The MAD estimation of the scale provided $\hat{\sigma}_x = 29.7$, and the tuning constant was chosen as $c = 2$.

Row 1 of Table 1 reports the benchmarks—namely, the LS-estimates produced by replacing outlying data with their conditional means. By comparing robust estimates, one sees that the M-method tends to overestimate θ_1 and to underestimate σ . This may be an indication of inefficiency. Finally, the second block provides the value of recursive estimates at ($t = T$).

Figure 2 shows the path of recursive estimates. All algorithms were initialized with $R(0) = I_4$, an identity matrix, and $\hat{\beta}(0) = \hat{\beta}_T$, the biased LS-estimates in row 2 of Table 1. The filter for the variances was (12) with $c = 2$ and initialized with previous robust estimates: $\hat{\sigma}(0) = 1$, $\hat{\sigma}_x(0) = 30$. From Figure 2, (a) and (b), one sees that recursive LS are much disturbed by outliers; however, in the case of system parameters $\delta_1, \omega_0, \omega_1$, they converge. The indications of Table 1 are confirmed in that M and GM-estimates differ significantly only for the noise parameter θ_1 . In particular, M-sequence does not converge to the benchmark in Table 1.

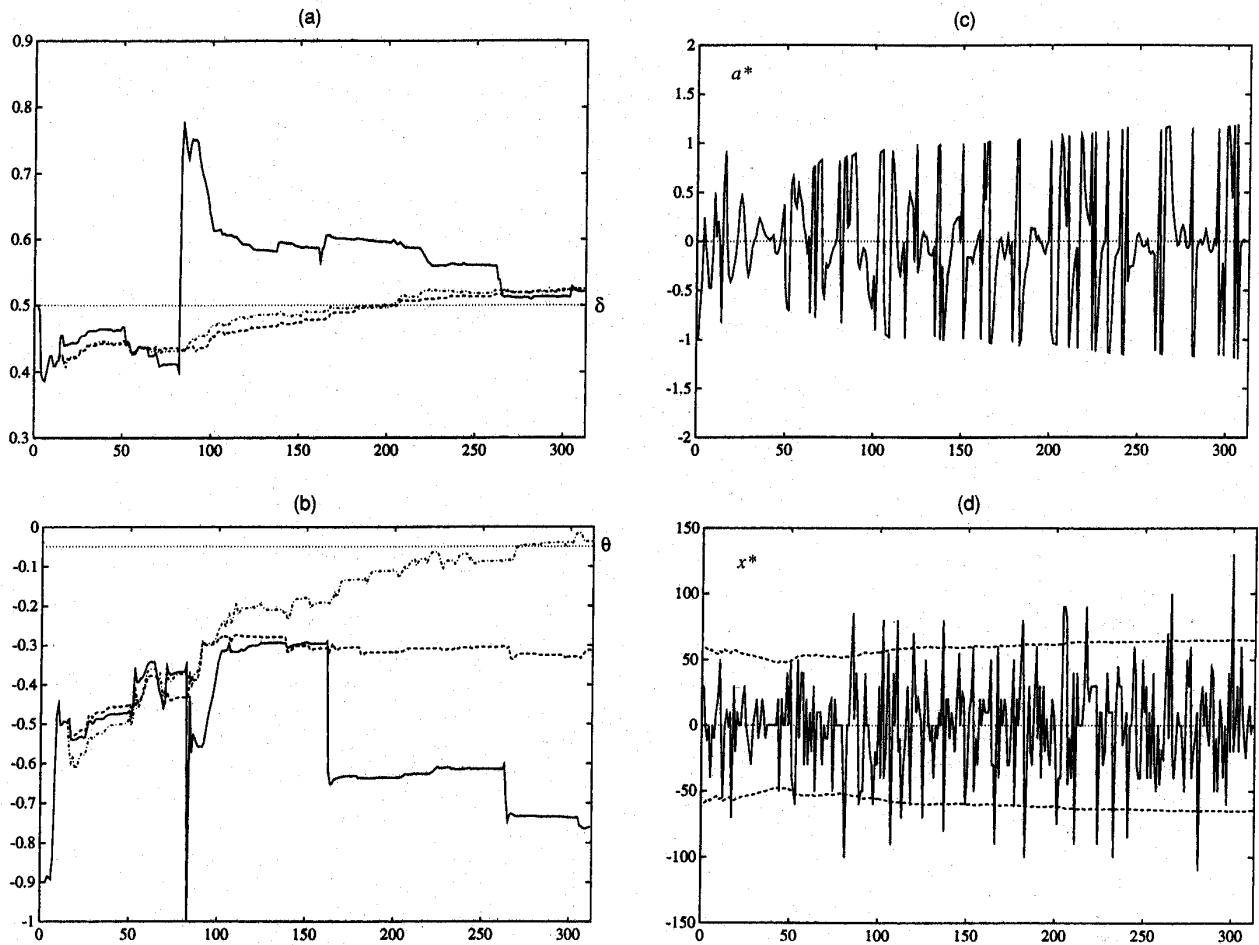


Figure 2. Recursive Estimates of Model (7) With LS, —, M, ---; GM, -.-. (a), (b): Parameters $\hat{\delta}(t), \hat{\theta}(t)$; (c), (d): series $\hat{a}^*(t), \hat{x}_t^*$ (—) with bands $\pm 2\hat{\sigma}(t)$ (-.-).

Table 2. Gas Furnace Data With Outliers Added: Parameter Estimates of Model (8)

Method	δ_1	ω_0	ϕ_1	θ_1	σ	ϕ_{x1}	ϕ_{x2}	σ_e
LS*	.695	-1.01	.782	.347	.142	1.69	-.765	.219
LS	.755	-.853	.727	-.516	1.21	.728	+.117	.712
M	.702	-1.01	.715	.023	.167	1.61	-.687	.216
GM	.692	-1.03	.732	.386	.171	1.68	-.772	.214
RLS	.749	-.823	.581	-.399	1.32	.897	-.051	.703
RM	.704	-1.00	.745	.187	.165	1.65	-.721	.176
RGM	.692	-1.04	.792	.387	.171	1.69	-.782	.201

NOTE: LS* = parameters estimated on data without AO's.

Because the results for ω_0, ω_1 are similar to those of δ_1 , they have been omitted. Instead, the cleaned series $\hat{a}^*(t)$ and \hat{x}_t^* of the GM-estimation have been displayed in Figure 2, (c) and (d), together with the bands $\pm 2\hat{\sigma}(t)$. The distinctive shape of $\hat{a}^*(t)$ may be attributed to the design $c = 2$, which is too restrictive because the outlier contamination is less than 5%.

Application 2. Regarding the gas furnace data with outliers added, the results of the iterative estimation of model (8) are given in Table 2. The first row reports the benchmarks—namely, the LS-estimates made without the AO's $y_{60}^o = -16, x_{110}^o = +6$. Robust algorithms were imple-

mented as in the previous application; however, the tuning constant was chosen as $c = 3$. We may see that GM is better than M in approaching the “true” values.

Figure 3 shows the recursive estimates of model (8). All algorithms were initialized with $\mathbf{R}(0) = 10 \cdot \mathbf{I}_4$ and $\hat{\beta}(0) = \hat{\beta}_T$, the values in row 1 of Table 2; robust ones used the scale (12) with $c = 3$. Dotted lines represent LS-estimates made in the absence of outliers and serve as benchmarks. As in Table 2, M- and GM-estimates are similar for the system parameters δ_1 , and ω_0 and differ significantly for θ_1 . Specifically, only the GM-method is resistant to outliers.

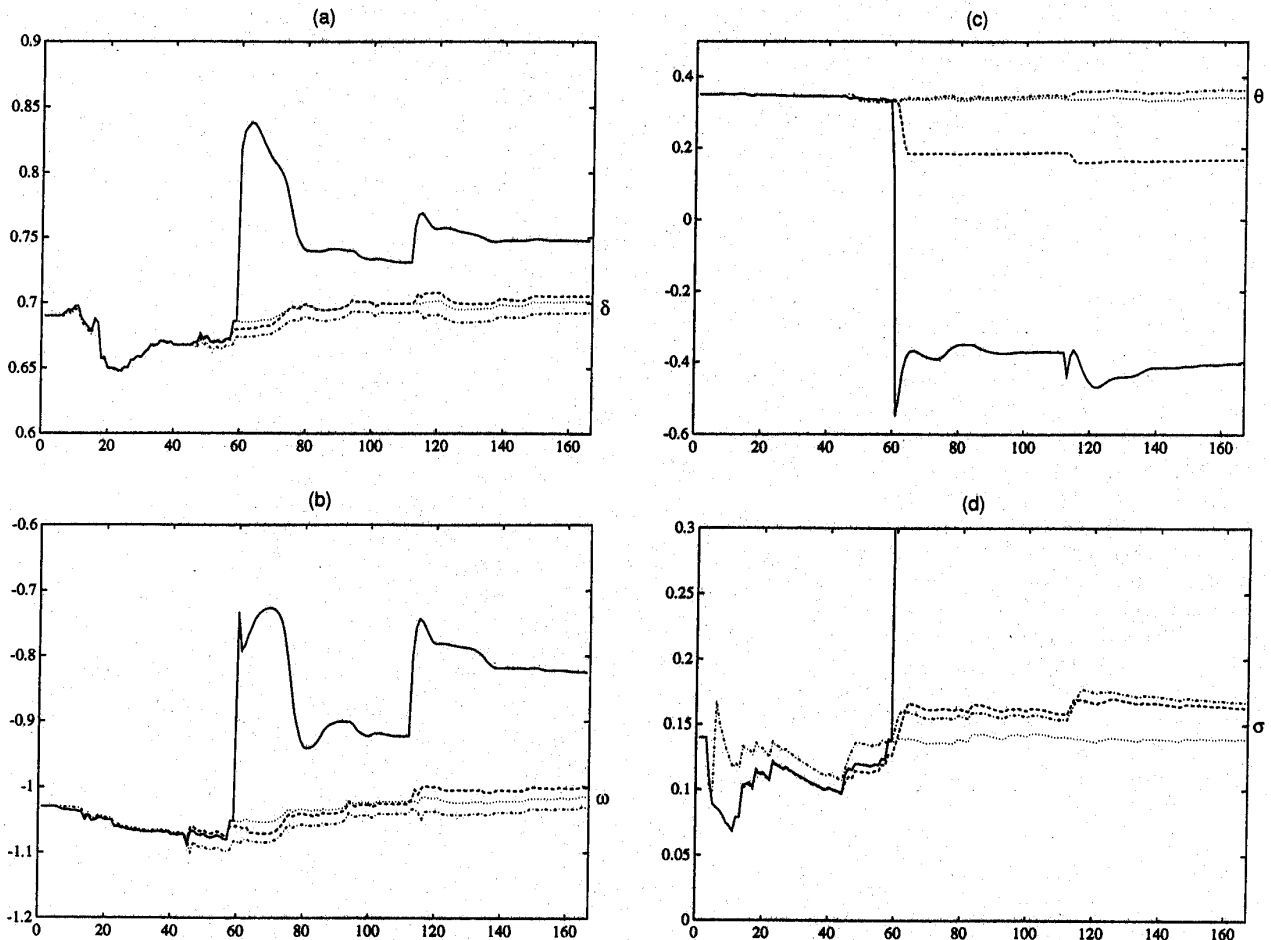


Figure 3. Recursive Estimates of Model (8) With LS, —; M, ---; GM, - · - ·. (a), (b): Parameters $\hat{\delta}_1(t), \hat{\omega}_0(t)$; (c), (d): parameters $\hat{\theta}_1(t), \hat{\sigma}(t)$ as in (12). LS* (···) denotes the estimates of the model without outliers added.

The conclusion of the preceding applications is twofold: (1) There is agreement between the results of iterative and recursive estimators, confirming the validity of the latter. (2) Only the GM-method is robust for each kind of parameter.

Asymptotic Properties. Although simulation results in Section 4 provide clear evidence of the robustness of Algorithm (15) in the presence of AO's, something must be said about its consistency in the case of $\{y_t, x_t\}$ without outliers. Under suitable regularity conditions, Ljung and Söderström (1983, p. 192) obtained the asymptotic distribution of the recursive M-estimator that minimizes the general loss function $P_t = \sum_{i=1}^t \rho(a_i)$. If $\rho(\cdot)$ is twice differentiable with respect to a_t , it is given by

$$\sqrt{t}[\hat{\beta}(t) - \beta] \xrightarrow{D} N[0, E(\xi_t \xi_t')^{-1} \kappa_\rho],$$

$$\kappa_\rho = \frac{E[\rho'(a_t)]^2}{\{E[\rho''(a_t)]\}^2}. \quad (16)$$

This result may be applied to algorithm (10), provided that $\psi_H(\cdot) = \rho'(\cdot)$ be made differentiable everywhere. As was discussed for (5), under the assumptions of stationarity of the inputs, stability of the system, and absence of outliers, the matrix $E(\xi_t \xi_t')$ in (16) does not depend on (t) .

The analysis of (15) is more involved because the vector $\{\xi_t^*\}$ does not, in general, equate the gradient. In iterative (off-line) form and in the absence of outliers, this algorithm actually coincides with the RA estimator, for which some asymptotic results were established by Bustos and Yohai (1986). In particular, in the case of ARMA models and for $\hat{\sigma}_T \rightarrow \sigma$ in probability, the distribution is given by (16) with κ_ρ replaced by

$$\kappa_\eta = \sigma^2 \frac{E[\eta^2(a_t/\sigma, a_{t-1}/\sigma)]}{\{E[\eta_1(a_t/\sigma, a_{t-1}/\sigma) a_{t-1}]\}^2}, \quad (17)$$

where $\eta_1(x, y) = \partial \eta(x, y) / \partial x$. This result may be extended to transfer function models and to recursive estimators because, for bounded initial conditions $\hat{\beta}(0)$, and $R(0)$, iterative and recursive implementations are asymptotically equivalent (see Ljung and Söderström 1983). The applicability of result (17) to Algorithm (15) is possible by taking the Mallows-type function $\eta(x, y) = \psi_H(x)\psi_H(y)$. Here, one may also define the relationships between (16) and (17) by noting that, for M-estimators, one has $\eta(x, y) = \psi(x)y$. Indeed, using this in κ_η , I may obtain κ_ρ with a_t just replaced by the rescaled residuals a_t/σ .

Apart from speed of calculation, an advantage of algorithm (15) is the relative simplicity of its dispersion matrix $E(\xi_t \xi_t')^{-1} \kappa_\eta$. With this, note that the efficiency of M and GM-estimators with respect to the LS-estimator is $1/\kappa$, independent of the system parameters β . Because κ is a function of η and therefore of ψ , it is obvious that the degree of efficiency crucially depends on the tuning constant c (see also Allende and Heiler 1992).

4. SIMULATION STUDIES

In this section I present simulation studies that were performed to check the robustness in the presence of AO's of the estimators developed in the article. Computations were carried out with the MATLAB package on a 486-33 personal computer.

Experiment 1. This experiment compares the performance of the iterative estimators developed in this article with the robust estimates presented by Allende and Heiler (1992).

I considered an ARMA(1, 1) model contaminated as in Equation (6b):

$$x_t = \phi x_{t-1} + \theta e_{t-1} + e_t, \quad e_t \sim \text{IN}(0, 1), \quad (18a)$$

and

$$x_t^o = x_t + u_t \nu_t, \quad u_t \sim \text{IB}(1, .05),$$

$$\nu_t \sim \text{IN}(0, 10\sigma_x), \quad (18b)$$

where e_t , u_t , and ν_t are mutually independent. In practice, the contamination is induced by a Gaussian white noise with variance 100 times greater than that of $\{x_t\}$ and with a proportion $\pi = 5\%$. From (18), $N = 100$ realizations of sample size $T = 100$ were generated independently. The tuning constant of the function $\psi_H(\cdot)$ was chosen as $c = 2$, which is consistent with the contamination proportion. Except for N , these designs are the same as the authors mentioned.

Part 1. First I consider $\phi = 0$ and $\theta = .8$ to also obtain a comparison with Bustos and Yohai (1986). Mean values and mean squared errors of the estimates are reported in Table 3. The first, two columns give the results as reported by Bustos and Yohai and Allende and Heiler for the methods LS, M, RA, and TRA (truncated RA). These were based on Huber's psi-function, on Mallows's eta-function, and on the MAD statistic as a measure of the scale. Instead, the last row provides the results of our GM-algorithm using two types of scale estimators. Note that the other authors did not supply any estimate of the scale σ .

By way of comment on Table 3, note that my algorithm based on parallel calculation performs significantly better than the RA and TRA ones, both in terms of mean values of the estimates and efficiency with respect to LS and M methods. Moreover, the performance of the algorithm does not depend on the scale measure. Note also that my LS and M estimates for θ are intermediate to those of the other authors. Because experimental designs are identical, their difference can be attributed to numerical factors, such as random-number generators, optimization programs, and computation algorithms. To compare robust estimates of different experiments, their efficiency with respect to the LS estimates is usually considered.

Part 2. I consider $\phi = .5$ and $\theta = .8$ for a second comparison with Allende and Heiler (1992). Their estimators were based on the Huber-Mallows functions and were implemented with the Hannan-Rissanen algorithm. They considered two scale measures—namely, MAD and Huber's proposal (in the last row). The results of the simulation are reported in Table 4.

By way of comment, one may note that GM-results of the two authors are nearly equivalent and quite satisfactory. As before, the difference between the LS-estimates may be attributed to numerical factors. Further experiments have

Table 3. Mean Values (and MSE) of the Estimates of Experiment 1 With $\phi = 0, \theta = .8, \sigma = 1$

Method	BY θ	AH θ	$\hat{\sigma}_e^2 = \text{MSE}$		$\hat{\sigma}_e \propto \text{MAD}$	
			θ	σ	θ	σ
LS	.08	.428 (.17)	.156 (.43)	2.96 (4.9)	.	.
M	.10	.443 (.15)	.304 (.28)	1.24 (.07)	.314 (.27)	1.20 (.07)
RA	.36	.526 (.09)
TRA	.40	.530 (.09)
GM	.	.	.623 (.04)	1.13 (.03)	.628 (.04)	1.12 (.03)

NOTE: BY = results of Bustos and Yohai (1986); AH = results of Allende and Heller (1992).

shown that the resistance to outliers of the GM-methods drastically improves by reducing the value of the parameter θ or the variance of the contamination process. We point out that the design conditions in this experiment are the most unfavorable ones considered by Allende and Heiler (1992).

Experiment 2. In testing recursive algorithms (10) and (14) the simulation conditions were redesigned as follows: (1) $N = 50$ replications of sample size $T = 200$ were considered. (2) Few outliers were introduced at fixed instants of time. This was done for monitoring the dynamic effects that outliers have on the path of recursive estimates.

The experiment deals with model (18a) contaminated by $x_{40}^o = 12, x_{80}^o = -16, x_{120}^o = 24, x_{160}^o = -16$. (19)

This choice was motivated by the fact that in the previous simulation I had $\max |x_t| \approx 8$. All algorithms were initialized with the conditions $\hat{\beta}(0) = 0, R(0) = I_2$; robust ones used the scale (12) with $c = 2$. Figure 4 displays the mean values $\bar{\beta}(t) = N^{-1} \sum_{i=1}^N \hat{\beta}_i(t)$ and the standard errors $SE_{\beta}(t) = \{N^{-1} \sum_{i=1}^N [\hat{\beta}_i(t) - \bar{\beta}(t)]^2\}^{1/2}$ of the recursive estimates.

Figure 4, (a) and (b), shows the time paths of $\bar{\phi}(t)$ and $\bar{\theta}(t)$; note that the performance of the GM-method is by far the most satisfactory. Figure 4, (c) and (d), provides the standard errors; it shows that the efficiency of recursive M and GM is similar. Finally, Figure 4, (e) and (f), displays the mean values of the scale estimates obtained with (12) and (13), and initialized with $\hat{\sigma}_e(0) = 1$. Because the MAD statistic is based on the median, it is natural that its time path is oscillatory, especially at the beginning. This feature is partially retained by the average $\bar{\sigma}(t) = N^{-1} \sum_{i=1}^N \hat{\sigma}_i(t)$.

Experiment 3. In this simulation I evaluated the performance of the iterative GM-estimator described in Appendix B on the simplified transfer function system

$$y_t = \frac{.5}{(1 - .8B)} x_{t-1} + a_t, \quad x_t \sim \text{IN}(0, 2^2), \quad a_t \sim \text{IN}(0, 1). \quad (20)$$

Both series $\{x_t, y_t\}$ were contaminated as in scheme (18b), and simulation conditions were the same as in Experiment 1—namely, $N = T = 100, c = 2$.

Numerical results are summarized in Table 5. Despite the high relative variance of the contaminating process and the fact that both input and output of the system were contaminated, note that the performance of the GM-estimator is quite satisfactory. This behavior was insensitive to the choice of the scale measure.

Experiment 4. This simulation is concerned with the recursive estimation of (20), contaminated by

$$x_{30}^o = y_{30}^o = 12, \quad x_{70}^o = y_{70}^o = -16, \quad x_{110}^o = y_{130}^o = 24, \quad x_{150}^o = y_{170}^o = -16. \quad (21)$$

This choice is motivated by the fact that in system (20) I had $\max |y_t| = \max |x_t| = 8$. The simulation conditions are the same as in Experiment 2—namely, $N = 50, T = 200$, and $c = 2$.

Numerical results are given in Figure 5. As before, Figure 5, (a) and (b), shows the mean values $\bar{\delta}(t), \bar{\omega}(t)$ of recursive estimates, and Figure 5, (c) and (d), displays the corresponding standard errors. Figure 5, (e) and (f), provides the paths of the mean values $\bar{\sigma}(t)$ obtained with (12) and (13). As in the previous case, the performance of the recursive GM-method is by far the most satisfactory. Results shown for parameter δ in Figure 5a are better than those shown for θ in Figure 4(b), although both make the system nonlinear in the parameters.

Experiment 5. Final simulation with iterative estimators considers the system

$$y_t = \frac{.5}{(1 - .5B)} x_{t-1} + \frac{(1 + .5B)}{(1 - .5B)} a_t, \quad x_t \sim \text{IN}(0, 2^2), \quad a_t \sim \text{IN}(0, 1), \quad (22)$$

where the contamination scheme for both $\{x_t, y_t\}$ is (18b). The simulation conditions are the same as in Experiments 1 and 3, and robust estimators are described in Appendix B.

Table 4. Mean Values (and MSE) of the Estimates of Experiment 1 With $\phi = .5, \theta = .8, \sigma = 1$

Method	Scale	Allende-Heiler		Proposed method		
		ϕ	θ	ϕ	θ	σ
LS	.	.158 (.14)	.143 (.47)	.469 (.15)	-.29 (1.4)	3.81 (9.4)
M	MAD	.177 (.12)	.149 (.46)	.658 (.04)	.201 (.39)	1.27 (.11)
GM	MAD	.466 (.02)	.665 (.04)	.502 (.03)	.657 (.06)	1.15 (.05)
GM	MSE	.387 (.04)	.875 (.02)	.501 (.02)	.621 (.07)	1.21 (.07)

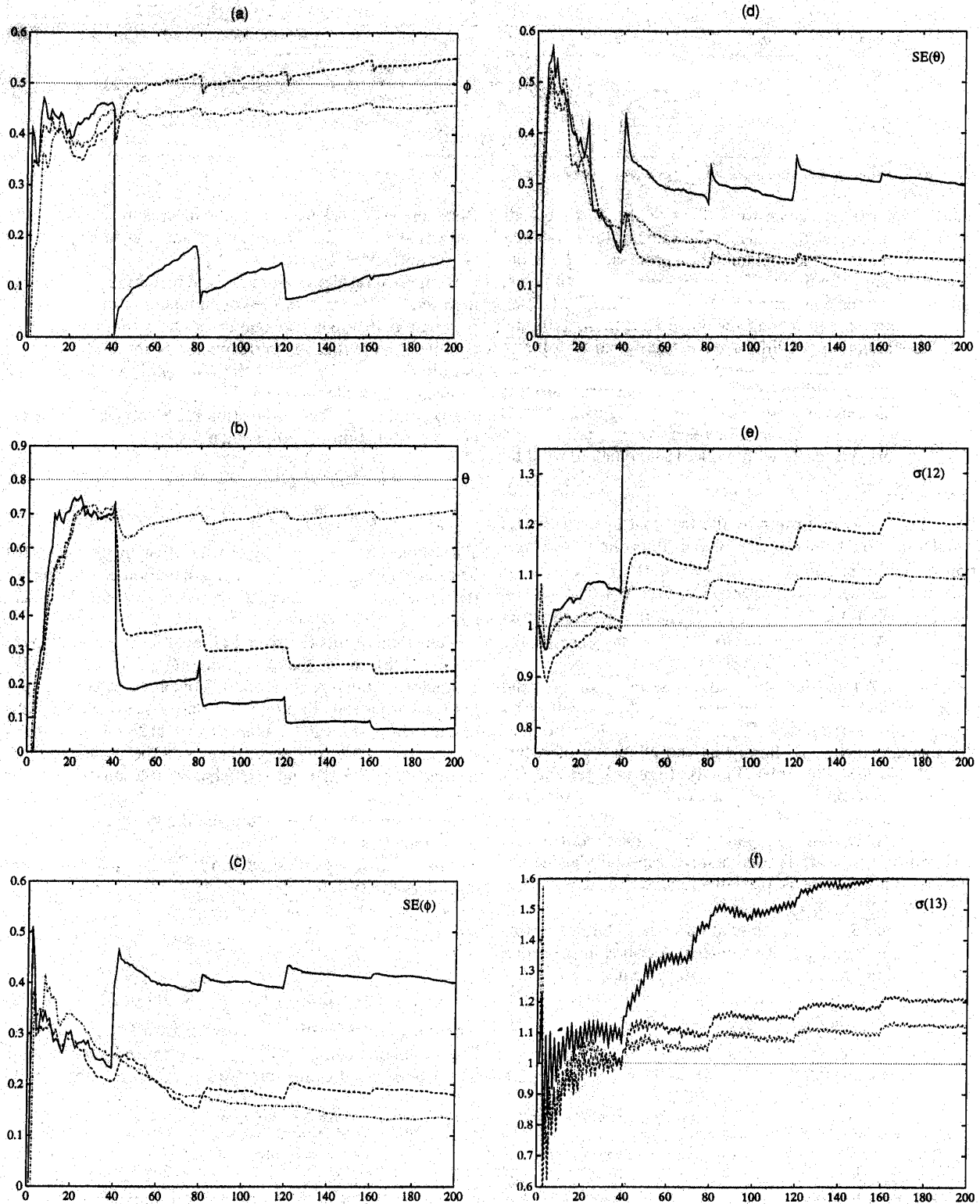


Figure 4. Results of Simulation Experiment 2 With LS, —; M, ---; GM, ····. (a) and (b) Mean values $\bar{\phi}(t)$, $\bar{\theta}(t)$; (c) and (d) standard errors; (e) and (f) mean values $\bar{\sigma}(t)$ of (12) and (13).

Table 5. Mean Values (and MSE) of Experiment 3 with $\delta = .8, \omega = .5, \sigma = 1, \sigma_x = 2$

Method	Scale	δ	ω	σ	σ_x
LS		.460 (.37)	.144 (.15)	3.98 (10.6)	4.56 (8.36)
M	MSE	.572 (.14)	.341 (.07)	1.61 (.43)	" "
M	MAD	.534 (.15)	.395 (.03)	1.52 (.35)	" "
GM	MSE	.783 (.003)	.456 (.008)	1.25 (.08)	2.07 (.03)
GM	MAD	.780 (.003)	.467 (.007)	1.22 (.07)	2.04 (.06)

Table 6. Mean Values (and MSE) of the Estimates of Experiment 5 With $\delta = \omega = \phi = \theta = .5$

Method	Scale	δ	ω	ϕ	θ	σ	σ_x
LS		.281 (.25)	.145 (.15)	.309 (.23)	-.13 (.59)	3.93 (9.8)	4.56 (8.4)
M	MSE	.512 (.05)	.374 (.06)	.521 (.06)	-.03 (.35)	1.42 (.23)	
M	MAD	.576 (.03)	.443 (.03)	.559 (.04)	+.06 (.22)	1.23 (.09)	
GM	MSE	.492 (.02)	.466 (.01)	.489 (.02)	.351 (.05)	1.19 (.05)	2.08 (.03)
GM	MAD	.493 (.02)	.472 (.01)	.481 (.02)	.383 (.04)	1.13 (.04)	2.04 (.06)

Table 7. Mean Values (and MSE) of the Estimates of Experiment 6 With $\delta = \omega = \phi = \theta = .5$

Method	δ	ω	ϕ	θ	σ	σ_x
LS	.476 (.04)	.201 (.09)	.443 (.04)	-.24 (.59)	2.96 (3.8)	3.20 (1.5)
M	.586 (.01)	.504 (.01)	.619 (.02)	.06 (.20)	1.15 (.03)	
GM	.483 (.01)	.509 (.01)	.494 (.01)	.460 (.01)	1.05 (.01)	1.97 (.01)
RLS ($t = T$)	.481	.200	.423	-.244	3.15	3.39
RM ($t = T$)	.535	.439	.579	.051	1.22	
RGM ($t = T$)	.468	.496	.485	.453	1.07	2.08

Numerical results are summarized in Table 6. As in the previous experiments, one may see that greatest differences between M and GM-methods concern the parameter θ . Moreover, the fact that GM-estimates of θ are not as good as for the other parameters may be attributed to the design conditions in (22), which are relatively unfavorable. Best performance is provided by the solution with the MAD measure for σ, σ_x .

Experiment 6. Final simulation with recursive estimators concerns System (22) contaminated as in (21). Experimental conditions are the same as Experiment 4; in particular, robust algorithms used the scale (12) with $c = 2$. Figure 6 reports the mean values of their estimates; one sees that the performance of the GM-method is satisfactory for each parameter.

To check the results in Figure 6, I also applied iterative algorithms to the same simulated data; the estimates are in Table 7. This enables one to have a precise comparison of the performance of on-line and off-line implementations. Note that Figure 6 and Table 7 do provide the same indications; in particular the GM-method is robust in the presence of AO's for every parameter, θ included.

5. CONCLUSIONS

In this article, I have developed robust recursive estimators for the parameters of dynamic models. The main feature is weighting and filtering data with factors that reduce outliers in the nonobservable regressors. Simulation experiments have shown that contamination by AO's has dramatic effects on LS and M-estimates; however, a substantial bias

reduction may be achieved with the GM-algorithm. Fundamental results are as follows:

1. For univariate models and iterative estimation, the proposed method is better than that of Bustos and Yohai (1986) and is at least as good as that of Allende and Heiler (1992). This situation legitimates the basic philosophy of the method.
2. For dynamic systems and recursive estimation, the proposed algorithms are at least as good as their iterative versions but are much faster. This remark encourages their use in practice, in particular in the control of industrial processes.

Further developments concern tracking of time-varying parameters. In this context, recursive algorithms must be made adaptive by discounting past observations. As shown by Grillenzoni (1994), robustification has a smoothing action on recursive adaptive estimates and may therefore be useful for systems that change slowly over time.

ACKNOWLEDGMENTS

I express my sincere gratitude to the editor, the associate editor, and two referees for the very constructive review process.

APPENDIX A: TECHNICAL NOTES

(1) *Winsorization.* This concept is related to that of *trimming* but differs in that weights assigned to outliers decrease more smoothly. In particular, although trimming simply discards observations that exceed a given threshold (c), Winsorization replaces them by the threshold value. In terms of psi-functions, the latter agrees with the Huber so-

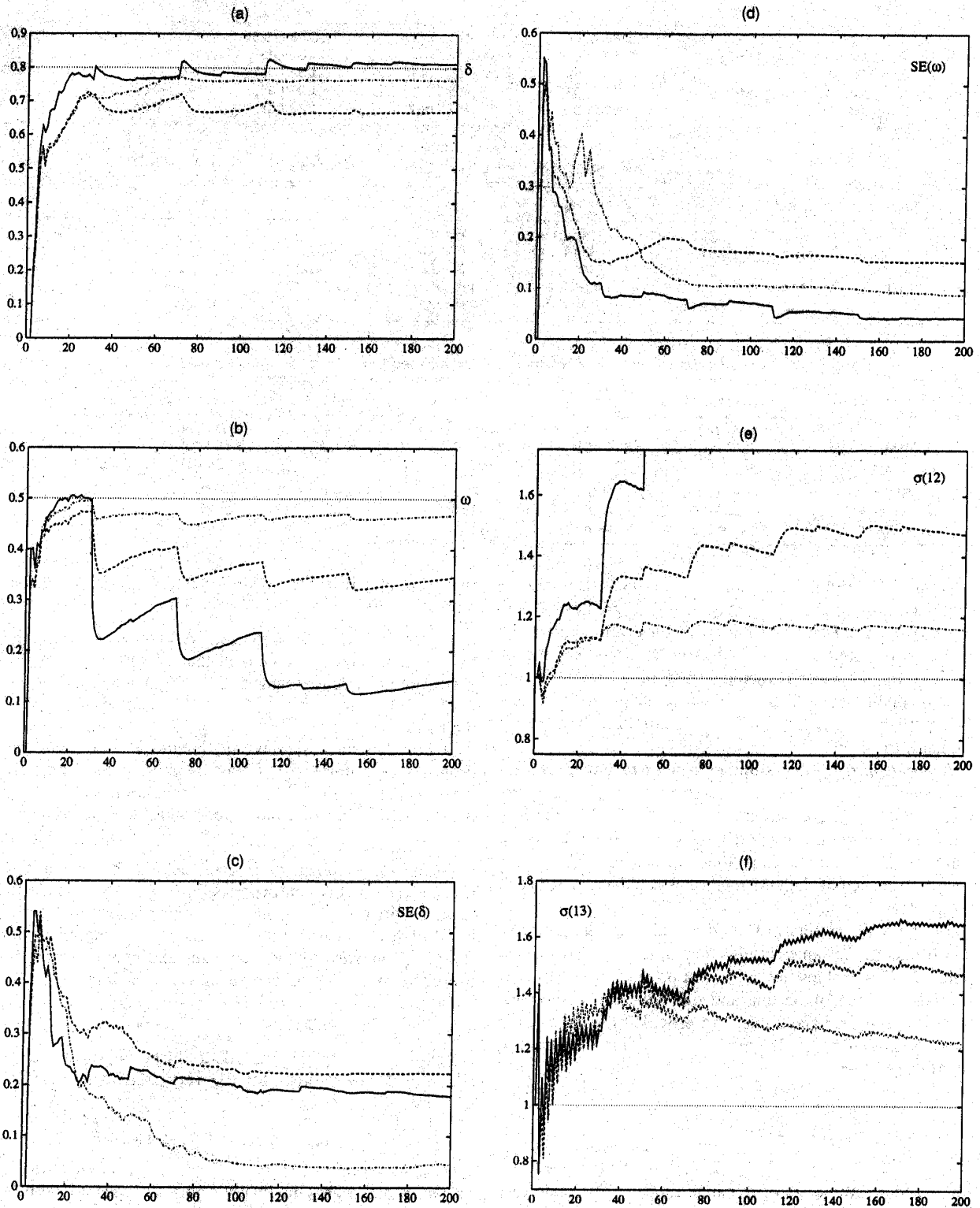


Figure 5. Results of Simulation Experiment 4 With LS, —; M, ---; GM, -.-. (a) and (b) Mean values $\bar{\delta}(t)$, $\bar{\omega}(t)$; (c) and (d) standard errors; (e) and (f) mean values $\bar{\sigma}(t)$ of (12) and (13).

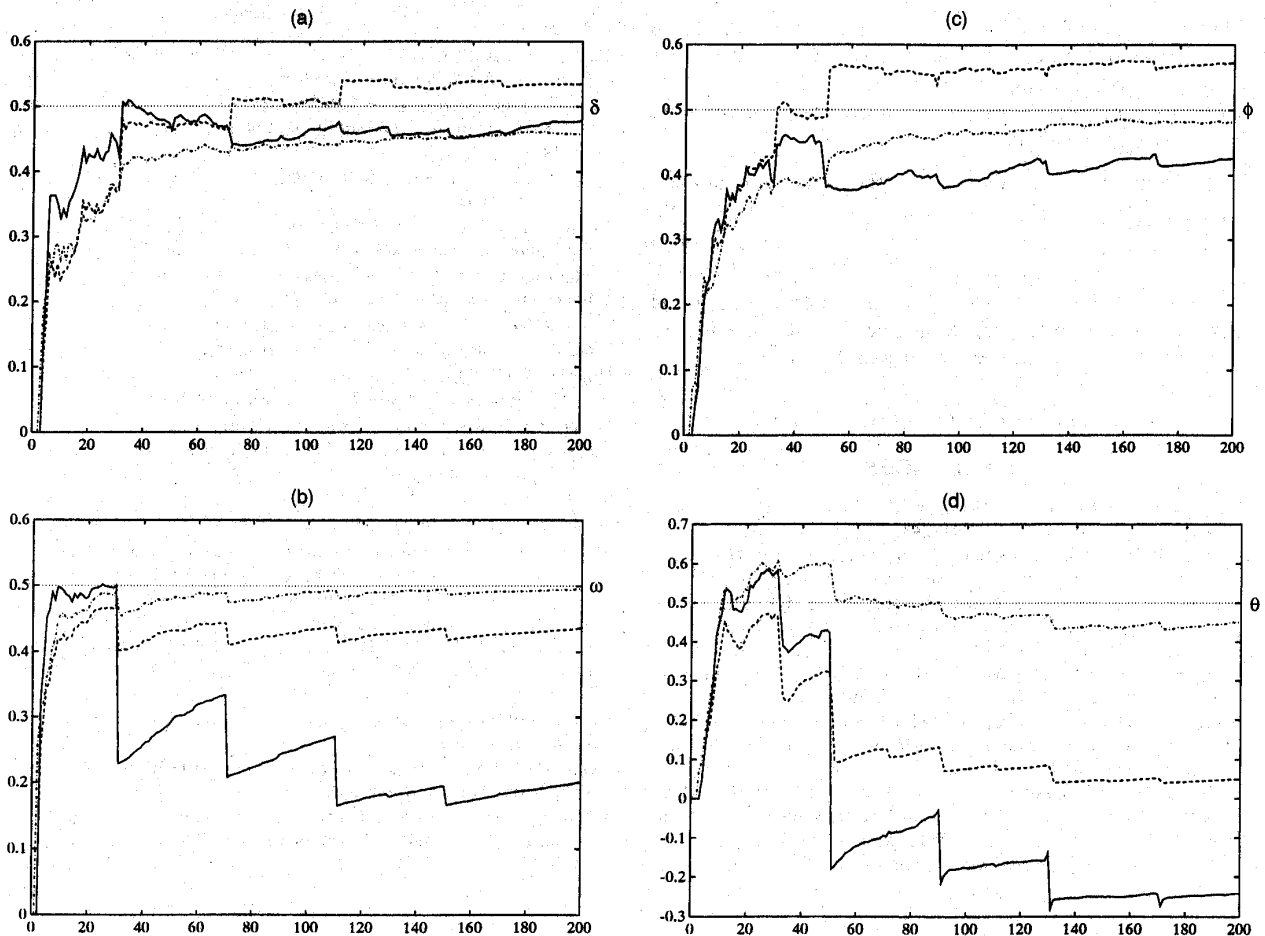


Figure 6. Results of Simulation Experiment 6 With LS, —; M, ---; GM, -.-. (a) and (b) Mean values $\bar{\delta}(t)$, $\bar{\omega}(t)$; (c) and (d) mean values $\bar{\phi}(t)$, $\bar{\theta}(t)$.

lution, whereas trimming corresponds to a function defined as $\psi_T(x) = 2x$ for $|x| \leq c$ and 0 elsewhere (Dudewicz and Mishra 1988, p. 740).

(2) *Bisquare*. The psi-function proposed by Beaton and Tukey (1974) is defined as $\psi_B(x) = x[1 - (x/c)^2]^2$ for $|x| \leq c$ and 0 elsewhere. This function is nonmonotone and discontinuous in $|x| = c$, and therefore its derivative $\psi'_B(x)$ may be negative when it exists. This feature raises serious problems to the Gauss-Newton estimator, which minimizes the functional $P_T = \sum_{t=1}^T \rho(a_t)$ and whose Hessian matrix is given by $R(t) = R(t-1) + \rho''(\hat{a}(t))\hat{\xi}(t)\hat{\xi}(t)'$ (see Ljung and Söderström 1983, p. 97). In fact, taking $\rho''(\cdot) = \psi'_B(\cdot)$, the matrix $R(t)$ would not be longer positive definite and convergence fails. For this reason, this article is only concerned with the Huber proposal.

(3) *BLUS Residuals*. As shown by Brown, Durbin, and Evans (1975), the recursive LS-estimator of σ is unbiased if it uses standardized prediction errors. In (10) these residuals are defined as $\bar{a}(t) = \hat{a}(t) \cdot [1/\hat{w}_c^2(t) + \hat{\xi}(t)'R(t-1)^{-1}\hat{\xi}(t)]^{-1/2}$ and are best linear unbiased and spherical (BLUS) for each $t > 0$ —namely, $(\bar{a}_t|y_{t-1}, x_{t-b}, y_{t-2}, \dots) \sim IN(0, \sigma^2)$. Now, having $R(t)^{-1} \rightarrow 0$ as $t \rightarrow \infty$, asymptotic unbiasedness is allowed even by nonstandard errors $\hat{a}(t)$.

APPENDIX B: AN ITERATIVE GM ALGORITHM FOR SYSTEM (1)

Because a preliminary robust estimator for the parameters of the model of the input series is needed, I rewrite Equation (1b) in regression form as $x_t = \alpha'x_t + e_t$, where $\alpha' = [\phi_{x1} \dots \phi_{xq}]$ and $x'_t = [x_{t-1} \dots e_{t-q}]$. Now, the iterative GM algorithm for system (1) consists of the following steps:

- (0) Get initial estimates $\hat{\alpha}_T, \hat{\beta}_T; \hat{\sigma}_e, \hat{\sigma}$ by applying LS to the contaminated series $\{x_t^o, y_t^o\}$.
- (1) Compute the residual of regression $\hat{e}_t = x_t - \hat{\alpha}'_T \hat{x}_t^*$ and clean it as $\hat{e}_t^* = \psi_H(\hat{e}_t/\hat{\sigma}_e)\hat{\sigma}_e$.
- (2) Compute the prediction $\hat{x}_t^* = \hat{\alpha}'_T \hat{x}_t^* + \hat{e}_t^*$ and update the vector $\hat{x}_{t+1}^* = [\hat{x}_t^* \dots \hat{e}_{t-q+1}^*]'$.
- (3) Repeat steps (1)–(2) for each $t = (p_x + 1), \dots, T$ so as to obtain the pseudo inputs $\{\hat{x}_t^*\}$; with this generate $\hat{m}_t^* = \sum_{i=1}^r \hat{\delta}_i \hat{m}_{t-i}^* + \sum_{j=0}^s \hat{\omega}_j \hat{x}_{t-b-j}^*$ by letting $\hat{m}_{-i}^* = 0$ for $i = 1 \dots r$.
- (4) Compute the residual $\hat{a}_t = y_t - \hat{\beta}'_T \hat{z}_t^*$ and clean it as $\hat{a}_t^* = \psi_H(\hat{a}_t/\hat{\sigma})\hat{\sigma}$.
- (5) Compute the prediction $\hat{y}_t^* = \hat{\beta}'_T \hat{z}_t^* + \hat{a}_t^*$ and the auxiliary variable $\hat{n}_t^* = \hat{y}_t^* - \hat{m}_t^*$; with these, update the vector of regressors $\hat{z}_{t+1}^* = [\hat{m}_t^* \dots \hat{x}_{t-b}^* \dots \hat{n}_t^* \dots \hat{a}_{t-p+1}^*]$.

(6) Repeat steps (4)–(5) for $t = (b + 1), \dots, T$ so as to obtain the pseudo outputs $\{\hat{y}_t^*\}$; next reestimate the parameters $\alpha, \beta; \sigma_e, \sigma$ by applying LS to the series $\{\hat{x}_t^*, \hat{y}_t^*\}$.

(7) Repeat steps (1)–(6) until the estimates of the noise variances σ_e^2, σ^2 converge.

From step (6) it is clear that the loss function implicit in this algorithm is $Q_T^*(\beta) = \sum_{t=1}^T (a_t^*)^2$. The method then proceeds by alternating the minimization of $\sum_t [\psi_H(a_t/\hat{\sigma}_T)\hat{\sigma}_T]^2$, with $a_t = y_t - \beta'z_t^*$, and the computation of $\hat{\sigma}_T^2 = T^{-1} \sum_{t=1}^T (\hat{a}_t^*)^2$. The corresponding M-estimator differs inasmuch as it does *not* use cleaned regressors z_t^* in the computation of residuals a_t .

[Received May 1994. Revised February 1996.]

REFERENCES

- Abraham, B., and Chuang, A. (1989), "Outlier Detection and Time Series Modeling," *Technometrics*, 31, 241–248.
- Allende, H., and Heiler, S. (1992), "Recursive Generalized M-Estimates for ARMA Models," *Journal of Time Series Analysis*, 13, 1–18.
- Beaton, A. E., and Tukey, J. E. (1974), "The Fitting of Power, Meaning Polynomials, Illustrated on Band-Spectroscopic Data," *Technometrics*, 16, 147–192.
- Box, G. E. P., and Jenkins, G. M. (1976), *Time Series Analysis: Forecasting and Control* (2nd ed.), San Francisco: Holden-Day.
- Brown, R. L., Durbin, J., and Evans, J. M. (1975), "Techniques for Testing the Constancy of Regression Relationships Over Time" (with discussion), *Journal of the Royal Statistical Society, Ser. B*, 37, 149–192.
- Bustos, O. H., and Yohai, V. J. (1986), "Robust Estimates for ARMA Models," *Journal of the American Statistical Association*, 81, 155–168.
- Cameron, M. A., and Turner, T. R. (1993), "Robust Recursive Location and Scale Estimators," *Communications in Statistics—Theory and Methods*, 22, 2503–2515.
- Chang, I., Tiao, G. C., and Chen, C. (1988), "Estimation of Time Series Parameters in the Presence of Outliers," *Technometrics*, 30, 193–204.
- Denby, L., and Martin, D. R. (1979), "Robust Estimation of the First-Order AR Parameter," *Journal of the American Statistical Association*, 74, 140–146.
- Dudewicz, J. N., and Mishra, S. R. (1988), *Modern Mathematical Statistics*, New York: Wiley.
- Gebski, V., and McNeil, D. (1984), "A Refined Method of Robust Smoothing," *Journal of the American Statistical Association*, 79, 616–623.
- Grillenzoni, C. (1991), "Iterative and Recursive Estimation of Transfer Functions," *Journal of Time Series Analysis*, 12, 123–144.
- (1994), "Optimal Recursive Estimation of Dynamic Models," *Journal of the American Statistical Association*, 89, 777–787.
- Hogg, R. V. (1977), "An Introduction to Robust Procedures," *Communications in Statistics, Part A—Theory and Methods*, 6, 813–827.
- Huber, P. J. (1981), *Robust Statistics*, New York: Wiley.
- Ljung, G. M. (1993), "On Outlier Detection in Time Series," *Journal of the Royal Statistical Society, Ser. B*, 55, 559–567.
- Ljung, L., and Söderström, T. (1983), *Theory and Practice of Recursive Identification*, Cambridge, MA: The MIT Press.
- Martin, R. D. (1980), "Robust Estimation of Autoregressive Models," in *Directions in Time Series*, eds. D. R. Brillinger and G. C. Tiao, Haywood, CA: Institute of Mathematical Statistics, pp. 228–254.
- Martin, R. D., and Yohai, V. J. (1985), "Robustness in Time Series and Estimating ARMA Models," in *Handbook of Statistics Volume 5*, eds. E. J. Hannan, P. R. Krishnaiah, and M. M. Rao, Amsterdam: North-Holland, pp. 119–155.
- Peña, D. (1991), "Measuring Influence in Dynamic Regression Models," *Technometrics*, 33, 93–101.
- Pierce, D. A. (1972), "Least Squares Estimation in Dynamic Disturbance Time Series Models," *Biometrika*, 59, 73–78.
- Schick, I. C., and Mitter, S. K. (1994), "Robust Recursive Estimation in the Presence of Heavy-Tailed Observation Noise," *The Annals of Statistics*, 22, 1045–1080.