The Annals of Statistics
1986, Vol. 14, No. 3, 781–818

INVITED PAPER

INFLUENCE FUNCTIONALS FOR TIME SERIES[1]

BY R. DOUGLAS MARTIN AND VICTOR J. YOHAI

University of Washington and University of Buenos Aires and CEMA, Buenos Aires

A definition is given for influence functionals of parameter estimates in time-series models. The definition involves the use of a contaminated observations process of the form $y_t^\gamma = (1 - z_t^\gamma)x_t + z_t^\gamma w_t$, $t = 1, 2, \dots$, $0 \le \gamma \le 1$, where x_t is a core process (usually Gaussian), w_t is a contaminating process, and z_t^γ is a zero–one process with $P(z_t^\gamma = 1) = \gamma + o(\gamma)$. This form is sufficiently general to model such diverse contamination types as isolated outliers and patches of outliers. Let $\mathbf{T}(\mu_y^\gamma)$ denote the functional representation of a given estimate, where the measures μ_y^γ, $0 \le \gamma \le 1$ for y_t^γ are in an appropriate subset of the family of stationary and ergodic measures on (R^∞, β^∞). The influence functional **IF** is a derivative of **T** along "arcs" traced by μ_y^γ as $\gamma \to 0$, and correspondingly $\mu_y^\gamma \to \mu_x$. Although this influence functional is similar in spirit to Hampel's influence curve **ICH** for the i.i.d. setting, it is not the same as **ICH**. However, a simple relationship between the **IF** and the **ICH** is established. Results are given which aid in the computation of **IF** and insure that **IF** is bounded. We compute the **IF** for some robust estimates of the first-order autoregressive and first-order moving average parameters using various contamination processes. A definition of gross-error sensitivity (**GES**) for the **IF** is given, and some estimates are compared in terms of their **GES**'s. Also the **IF** is used to show that a class of generalized RA estimates has a certain optimality property. Finally, some possible generalizations of the **IF** are indicated.

1. Introduction. The influence curve, introduced by Hampel (1974), has been referred to as "perhaps the most useful heuristic tool of robust statistics" by Huber (1981) in Chapter 1.5 of his recent book. Indeed the usefulness of the influence curve in situations where the data consist of independent and identically distributed (i.i.d.) random variables or random vectors is reflected by its appearance in many papers on robustness, and by attempts to extend its definition to cover situations other than the standard point estimation problems. For example one finds recent papers on influence curves in the context of errors in variables (Kelly, 1984), quantal bioassay (James and James, 1983), problems involving censoring (Samuels, 1978), and for parameter testing (Lambert, 1981; Ronchetti, 1982), and goodness of fit tests (Michael and Schucany, 1985).

In spite of the pervasive nature of the influence curve and the length of time elapsed since Hampel's initial contribution, a completely satisfactory definition of

Received July 1984; revised April 1986.

[1] Research supported by the Office of Naval Research under contracts N00014-82-K-0062 and N00014-84-C-0169.

AMS 1980 *subject classifications.* Primary 62G35; secondary 62M10, 62F10.

Key words and phrases. Influence functionals, influence curves, time series, robust estimates.

influence curve for the time-series setting has not yet been given. A number of authors have suggested carrying over Hampel's definition of influence curve to the time-series setting: Martin and Jong (1976), Portnoy (1977), and Martin (1980) mention this possibility, while Künsch (1984) has pursued the use of Hampel's influence curve for obtaining infinitesimal optimality results for autoregression estimates. See also Chernick, Downing, and Pike (1982). However, we shall argue that while Hampel's influence curve plays a central role in reflecting the "influence" of contamination in the i.i.d. setting, it does not adequately capture the nature of contamination effects in the time-series setting. We attempt to remedy the situation by providing a useful definition of influence functional (**IF**) for time-series parameter estimation problems which seems natural and closely coupled to intuition.

One of the chief features of the time-series setting is the fact that estimators which take account of the time-series structure are *not* invariant under permutations of the data, as in the case of estimators for i.i.d. situations. Consequently, basic permutation dependent issues of contamination, such as the distinction between outliers which occur in isolation versus outliers which occur in patches, become important. Such distinct types of behavior are common occurrences in real data, as any careful and experienced practitioner knows all too well. Our definition of influence functional reflects the difference in impact of these two types of behavior, as well it should. These differences are clearly revealed in some explicit computations of influence functionals.

As a point of departure we briefly recall Hampel's (1974) definition of influence curve and its properties. The context is that of possibly vector-valued *independent* observations $\mathbf{y}_1, \dots, \mathbf{y}_n$ with common distribution F, and an estimator $\mathbf{T}_n = \mathbf{T}_n(\mathbf{y}_1, \dots, \mathbf{y}_n)$ which may also be vector-valued. It is assumed that $\mathbf{T}_n$ may be obtained from a functional $\mathbf{T} = \mathbf{T}(F)$ defined on a suitably rich family of distributions by evaluating $\mathbf{T}$ at the empirical distribution function F_n: $\mathbf{T}_n = \mathbf{T}(F_n)$. Let $F_\gamma = (1 - \gamma)F + \gamma\delta_{\mathbf{y}}$ be a contamination distribution, where $\delta_{\mathbf{y}}$ has all its mass at $\mathbf{y}$. Then Hampel's influence curve is the directional or Gateaux derivative at F of the functional $\mathbf{T}$, in the "direction" determined by $\delta_{\mathbf{y}}$:

$$\mathbf{ICH}(\mathbf{y}) = \mathbf{ICH}(\mathbf{y}; \mathbf{T}, F) = \lim_{\gamma \to 0} \frac{\mathbf{T}(F_\gamma) - \mathbf{T}(F)}{\gamma} \tag{1.1}$$

The influence curve is both an asymptotic and a local (or infinitesimal) tool.

The influence curve has several useful properties (Hampel, 1974):

(P1) an appealing heuristic interpretation;
(P2) a convenient role in formal asymptotics;
(P3) an indicator via gross-error sensitivity (GES) of maximum bias due to infinitesimal contamination;
(P4) the construction of optimal estimates under the constraint of a bounded gross-error sensitivity (GES).

Results on (P4) for ordinary regression may be found in Hampel (1975, 1978), Krasker and Welsch (1982), and Huber (1983). Similar results based on **ICH** for autoregression have been obtained by Künsch (1984).

The remaining parts of the paper are as follows: As preliminaries, Section 2 introduces a general class of contamination processes which is useful for our definition of influence functional (**IF**), along with several particular types of contamination, and, also, some notation used in the remainder of the paper. Section 3 discusses functional representations for time-series parameter estimates, introduces $\tilde{\Psi}$ estimates, the main class with which we work, and gives some specific $\tilde{\Psi}$ estimates.

Section 4 introduces our definition of **IF**, gives results which aid in the computation of the **IF**, and which insure boundedness of the **IF**. Section 5 gives specific results for generalized M-estimates and RA estimates of first-order autoregressive and moving average models. The results given in Sections 4 and 5 address Künsch's (1984, Section 2.6) second open question in the context of our definition of time series **IF**: namely, we include the case where the estimator depends upon the measure for the process (not just on finite-dimensional marginal measures). In particular, it is shown that although bounded psi functions (i.e., bounded summands in estimating equations) yield a bounded **IF** for AR(1) models, this condition is not sufficient to insure a bounded **IF** for MA(1) models. On the other hand, redescending psi functions can yield a bounded **IF** for MA(1) models. These results revel a key distinction with regard to robustness between models having moving average components and those which do not.

Section 6 introduces a definition of gross-error sensitivity (GES), and computes GES's for the estimates and models treated in Section 5. Section 7 introduces a class of generalized RA-estimates and establishes, using the definition of **IF**, a certain optimality of these estimates. Section 8 sketches some possible generalizations of the **IF** through applications to a white-noise test statistic, and to spectral density estimation. Finally, proofs of theorems are collected in Section 9.

Throughout, we keep (P1)–(P4) of the **ICH** in mind, with a view toward preserving the most essential of these in the time-series setting. Our preference ranking for these properties makes (P1) and (P3) paramount, with (P4) highly desirable.

2. Contamination processes for time series.

2.1. *The importance of outliers' time configurations.* In the case of estimates which are intended for use in the i.i.d. setting, such as ordinary location M estimates, the influence curve may be defined asymptotically as in (1.1), with Hampel's (1974) attendant finite-sample size approximation, by placing all the contamination at a single point. This approach works essentially because most estimators intended for use with i.i.d. data are invariant under permutations of the data, and may be obtained from functionals **T** of the marginal distribution function F by evaluating **T** at the empirical marginal distribution function F_n. Under such conditions, the specific *time configuration* of observation times at which the contaminating points occur is irrelevant.

By way of contrast, the time configuration of the contaminating points will be important in the case of estimates which make use of the time-series structure.

For the sake of specific illustration consider the ordinary lag-one correlation coefficient

$$\hat{\rho} = \frac{\sum_{i=2}^{n} y_i y_{i-1}}{\sum_{i=1}^{n} y_i^2}$$

with estimation of the mean ignored for simplicity (the behavior to be described is qualitatively the same when $\bar{y}$ is included). It is clear that the values of a fixed number k of *isolated* outliers appear quadratically in the denominator of $\hat{\rho}$, but only linearly in the numerator; by "isolated" we mean that each pair of outliers is separated in time by at least one nonoutlier observation. If the outliers have a common value or "amplitude" ζ, then $\hat{\rho} \to 0$ as $\zeta \to \infty$, and the effect might be described as bias toward zero. On the other hand, if the k outliers of common amplitude ζ are *contiguous*, i.e., if they form a *patch* of length k, then $\hat{\rho} \to (k-1)/k$ as $\zeta \to \infty$. For long patch lengths k, the effect might be described as bias toward unity.

Since different time configurations can have quite different impacts on an estimate, it will be natural when defining an influence functional for time series to work with contaminated processes which have the flexibility to provide different time configurations of contamination or outliers, as well as a controlled contamination fraction.

2.2. *The general replacement model.* The following component processes and associated stationary and ergodic marginal measures on $(R^\infty, \mathscr{B}^\infty)$ are used to construct the contaminated process:

$x_i \sim \mu_x$, the nominal or core process, often Gaussian,

$w_i \sim \mu_w$, a contaminating process,

$z_i^\gamma \sim \mu_z^\gamma$, a 0–1 process,

where $0 \le \gamma \le 1$, and

$$P(z_i^\gamma = 1) = g(\gamma) = \gamma + o(\gamma) \tag{2.1}$$

for some function g. The contaminated process y_i^γ is now obtained by the general *replacement* model

$$y_i^\gamma = (1 - z_i^\gamma)x_i + z_i^\gamma w_i, \tag{2.2}$$

where $y_i^\gamma \sim \mu_y^\gamma$ and $\mu_y^0 = \mu_x$, i.e., zero contamination results in perfect observations of x_i. In general we may wish to allow dependence between the z_i^γ, x_i, and w_i processes in order to model certain kinds of outliers. Correspondingly, the measures μ_y^γ, $0 \le \gamma \le 1$, are determined by the specification of the joint measures μ_{xwz}^γ, $0 \le \gamma \le 1$.

The pure replacement model. Here z_i^γ, x_i, and w_i are mutually independent processes, i.e.,

$$\mu_{xwz}^\gamma = \mu_z^\gamma \mu_x \mu_w.$$

The AO model. Allowing dependence between x_t and w_t means that the additive outliers (AO) model

$$y_t = x_t + v_t^* \tag{2.3}$$

used in previous studies (Denby and Martin, 1979; Bustos and Yohai, 1986) may in some situations be obtained as a special case of (2.2). This is the case for example when the v_i^* have, for marginal distribution, the contamination distribution $F_v = (1-\gamma)\delta_0 + \gamma H$ with degenerate central component δ_0. Just set $w_t = x_t + v_i$ with v_i having marginal distribution H, let $g(\gamma) = \gamma$ in (2.2), and let z_i^γ be independent of x_i and v_i. Here $\mu_{xwz}^\gamma = \mu_z^\gamma \mu_{xw} = \mu_z^\gamma \mu_{x,x+v}$. Throughout the remainder of the paper we use the version of the AO model obtained from (2.2) with $w_t = x_i + v_t$.

The two main time configurations for outliers are (a) isolated outliers, and (b) outliers occurring in patches or bursts. The need for modeling the latter behavior is well recognized by those who have dealt with real time-series problems. It may also be desirable to combine these two situations in order to adequately model some time-series data.

Independent outliers. Since isolated outliers are typically produced by independence in the w_t or v_t, we shall use the terms "independent" and "isolated" as interchangeable adjectives.

Situations in which the outliers are mainly isolated are easily manufactured from either the pure replacement or AO form of (2.2) by letting z_i^γ be i.i.d. with $g(\gamma) = \gamma$ and w_t an appropriately specified process. For example, w_i could be an i.i.d. Gaussian process with mean zero and suitably large variance, or the w_i could be identically equal to a constant value ζ.

Patchy outliers. Patches of approximately fixed length can be arranged in the following way. Let w_i and v_t be highly correlated processes. In case these processes are identically equal to a constant ζ, they will be regarded as highly correlated. Now let $\tilde{z}_i^p$ be an i.i.d. binomial $B(1, p)$ sequence, and set

$$z_i^\gamma = \begin{cases} 1, & \text{if } \tilde{z}_{t-l}^p = 1 \text{ for some } l = 0, 1, \dots, k-1, \\ 0, & \text{else.} \end{cases} \tag{2.4}$$

Here we set $\gamma = kp$, with k fixed and p variable. Then since

$$P(z_i^\gamma = 1) = 1 - (1-p)^k = kp + o(p)$$

we have

$$g(\gamma) = \gamma + o(\gamma) \tag{2.5}$$

and the average patch length is k for γ small. We denote the probability measure of the process $\{z_t^\gamma\}$ by $\mu_z^{k,\gamma}$.

2.3. *Some notation.* In the sequel we shall use the following notational conventions. Finite sets of contiguous x_i will be denoted

$$\mathbf{x}_i^j = (x_i, x_{t-1}, \dots, x_j), \qquad j \le i, \tag{2.6}$$

and similarly with w_i. We often need dummy arguments which are shadow representatives for observations, and we use y_i's for this purpose. Correspondingly, finite segments of y_i's are denoted

$$\mathbf{y}_i^j = (y_i, y_{i-1}, \dots, y_j), \qquad j \le i. \tag{2.7}$$

We will have little need to refer to finite segments of y_i^γ and z_i^γ. However, we do need semiinfinite sequences of y_i^γ's (with measure μ_y^γ) and x_i's (with measure μ_x), as well as semiinfinite sequences of dummy arguments y_i. These we denote by

$$\mathbf{y}_i^\gamma = (y_i^\gamma, y_{i-1}^\gamma, \dots), \tag{2.8}$$

$$\mathbf{x}_i = (x_i, x_{i-1}, \dots), \tag{2.9}$$

$$\mathbf{y}_i = (y_i, y_{i-1}, \dots), \tag{2.10}$$

and so on. The use of $\mathbf{y}_i^\gamma$ and $\mathbf{x}_i$ is almost exclusively reserved for computations involving expectations under μ_y^γ and μ_x, respectively, and since these are stationary measures, our typical arguments are $\mathbf{y}_1^\gamma$ and $\mathbf{x}_1$. Since $\mathbf{y}_i$ is usually a dummy argument, we usually use $\mathbf{y}_1$.

Semi-infinite sequences such as (2.8)–(2.9) may be regarded as points in R^∞. Doubly infinite sequences such as $(\dots, y_{-1}, y_0, y_1, y_2, \dots)$ are points in the space $R^{-\infty,\infty}$ of all doubly infinite sequences.

3. Time-series parameter estimates and functionals. It is assumed throughout that the observations y_i are realizations of a stationary and ergodic process on $R^{-\infty,\infty}$, with associated probability space $(R^{-\infty,\infty}, \mathscr{B}, \mu)$, $\mathscr{B}$ being the family of Borel sets in $R^{-\infty,\infty}$, with μ in the set $\mathbf{P}_{se}$ of all stationary and ergodic measures on $(R^{-\infty,\infty}, \mathscr{B})$. In this time-series setting, it is usually possible to represent the asymptotic value of a parameter estimate as a functional, $\mathbf{T} = \mathbf{T}(\mu)$, defined on a subset $\mathbf{P}_0$ of $\mathbf{P}_{se}$.

The basic definitions of influence functional and gross-error sensitivity for time-series parameter estimates, which we give subsequently, are for quite general functionals $\mathbf{T}(\mu)$. However, all specific ensuing results are for a special class of functionals $\mathbf{T}$ associated with those time-series models parameter estimates $\mathbf{T}_n$ which may be computed as a solution to the estimating equation

$$\sum_{i=1}^{n} \tilde{\mathbf{\Psi}}_i(y_i, \dots, y_1, \mathbf{T}_n) = \sum_{i=1}^{n} \tilde{\mathbf{\Psi}}_i(\mathbf{y}_i^1, \mathbf{T}_n) = \mathbf{0}. \tag{3.1}$$

Here each $\tilde{\mathbf{\Psi}}_i$ is a function from $R^i \times R^m$ to R^m. Both $\tilde{\mathbf{\Psi}}_i$ and $\mathbf{T}_n$ may be vector-valued, as when estimating the parameters of an autoregressive-moving-average model of orders p and q. For the sake of notational simplicity we take the y_i to be scalar-valued, but all of what follows applies equally well to the case of vector-valued time series.

The subscript i on $\tilde{\mathbf{\Psi}}_i$ accounts for "end effects" which vanish either after a finite number of observations (as in Example 1 to follow), or asymptotically (as in Examples 2–4 to follow). In either case the asymptotic value $\mathbf{T} = \mathbf{T}(\mu)$ of $\mathbf{T}_n$ can usually be determined through the use of a fixed psi function $\tilde{\mathbf{\Psi}}$ which for each $\mathbf{t}$

satisfies

$$(3.2)\qquad \lim_{i\to\infty} \tilde{\Psi}_i(a_1,\dots,a_i,\mathbf{t}) = \lim_{i\to\infty} \tilde{\Psi}(\mathbf{a},\mathbf{t}), \qquad \forall \mathbf{a} = (a_1, a_2, \dots) \in R^\infty.$$

A special case of such a $\tilde{\Psi}$ is one which depends only on a finite number of coordinates: for each $\mathbf{t}$

$$(3.2')\qquad \tilde{\Psi}_i(a_1,\dots,a_i,\mathbf{t}) = \tilde{\Psi}(\mathbf{a},\mathbf{t}), \qquad i \ge k, \forall \mathbf{a} \in R^\infty.$$

Example 1 to follow falls into this category, whereas Examples 2–4 require the general form of $\tilde{\Psi}$.

Under suitable regularity conditions, which include ergodicity, one expects to have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \tilde{\Psi}_i(y_i,\dots,y_1,\mathbf{t}) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \tilde{\Psi}(\mathbf{y}_i,\mathbf{t}) = E\tilde{\Psi}(\mathbf{y}_1,\mathbf{t}).$$

Therefore we assume that the asymptotic value $\mathbf{T}(\mu)$ is defined by

$$(3.3)\qquad \int \tilde{\Psi}(\mathbf{y}_1,\mathbf{T})\, d\mu(\mathbf{y}_1) = 0.$$

We shall assume that (3.3) either has a unique root $\mathbf{t}_0 = \mathbf{T}(\mu)$, or that a well-defined solution is available in the case of multiple roots. $\mathbf{T}$ is then defined on $\mathbf{P}_0$ consisting of all μ in $\mathbf{P}_{se}$ such that the integral in (3.3) exists and is finite.

An estimate $\mathbf{T}_n$ defined by (3.1) is called a $\tilde{\Psi}$ *estimate*, and this term will also be used to describe the associated asymptotic version $\mathbf{T}$ defined by (3.3). The class of $\tilde{\Psi}$ estimates is quite large and contains both classical and robust parameter estimates, as the examples to follow show.

Our examples consist of two classes of robust estimates of the parameters of first-order autoregressive, AR(1), and first-order moving-average, MA(1), models. The AR(1) model is

$$(3.4)\qquad x_i = \phi x_{i-1} + u_i$$

and the MA(1) model is

$$(3.5)\qquad x_i = u_i - \theta u_{i-1},$$

where the *innovations* u_i are assumed to be i.i.d. with a common $N(0,1)$ distribution. The assumption of a known innovations scale $\sigma_u = 1$ and known location $\mu = 0$ is made to simplify the exposition.

EXAMPLE 1. *GM/BIF estimates for the AR*(1) *model.* Here $\mathbf{T}_n = \hat{\phi}$ is a *generalized M estimate* (GM estimate) or *bounded-influence estimate* (BIF estimate) of ϕ. These estimates are $\tilde{\Psi}$ estimates, with $\tilde{\Psi}_i = \tilde{\psi}_i$ given by

$$(3.6)\qquad \tilde{\psi}_i(\mathbf{y}_i^1, \phi) = \eta\big(y_i - \phi y_{i-1}, y_{i-1}(1-\phi^2)^{1/2}\big), \qquad i \ge 2,$$

for some bounded function $\eta = \eta(\cdot,\cdot)$. Correspondingly, the limit $\tilde{\Psi}$ function is

$$(3.7)\qquad \tilde{\psi}(\mathbf{y}_1, \phi) = \eta\big(y_1 - \phi y_0, y_0(1-\phi^2)^{1/2}\big).$$

The two main variants of GM/BIF estimates are as follows (see Denby and Martin, 1979; Martin, 1980; Bustos, 1982).

Mallows variant.

$$\eta(\xi_1, \xi_2) = \psi(\xi_1)\psi(\xi_2) \tag{3.8}$$

for some bounded robustifying psi function ψ. This type of estimate was suggested by Mallows (1976) in the non-time-series regression setup.

Hampel–Krasker–Welsch (HKW) variant.

$$\eta(\xi_1, \xi_2) = \psi(\xi_1\xi_2) \tag{3.9}$$

for some bounded robustifying function ψ.

The choice $\eta(\xi_1, \xi_2) = \xi_2\psi(\xi_1)$ yields the ordinary M estimate of ϕ (see for example Martin, 1982). As we shall see, this estimate does not have a bounded influence function, using either our definition or Hampel's definition as extended to the time series context (see Section 5). The ordinary M estimate and the Mallows and HKW type estimates all reduce to the least-squares estimate, which has an unbounded influence function, by either definition when ψ is the identity function.

EXAMPLE 2. *RA estimates for the AR*(1) *model.* Recently Bustos and Yohai (1986) have introduced a new class of estimates for ARMA models. These estimates are called *RA estimates* because they are based on robust estimates of *residual autocovariances*. For the AR(1) case the RA estimate $\hat{\phi}$ is defined as follows. Let

$$r_t(\hat{\phi}) = y_t - \hat{\phi}y_{t-1} \tag{3.10}$$

and let

$$\hat{\gamma}_l = \gamma_l(\hat{\phi}) = \frac{1}{n}\sum_{t=l+2}^{n} \eta(r_t(\hat{\phi}), r_{i-l}(\hat{\phi})), \qquad 0 \le l \le n-2, \tag{3.11}$$

denote a robust lag-l autocovariance estimate for the residuals, with robustifying function $\eta = \eta(\cdot, \cdot)$. Then $\hat{\phi}$ is a solution of the estimating equation

$$\sum_{l=1}^{n-2} \hat{\phi}^{l-1}\gamma_l(\hat{\phi}) = \frac{1}{n}\sum_{t=3}^{n}\sum_{l=1}^{i-2} \hat{\phi}^{l-1}\eta(r_i(\hat{\phi}), r_{t-l}(\hat{\phi})) = 0. \tag{3.12}$$

Estimates obtained for the choices $\eta(\xi_1, \xi_2) = \psi(\xi_1)\psi(\xi_2)$ and $\eta(\xi_1, \xi_2) = \psi(\xi_1\xi_2)$ are again called Mallows and HKW estimates, respectively. It may be shown that if $\eta(\xi_1, \xi_2) = \xi_1\xi_2$, then $\hat{\phi}$ is asymptotically equivalent to the least-squares estimate, and if $\eta(\xi_1, \xi_2) = \psi(\xi_1)\xi_2$, then $\hat{\phi}$ is asymptotically equivalent to an M estimate.

Again $\tilde{\boldsymbol{\Psi}}_t = \tilde{\psi}_i$ is scalar-valued, and $\tilde{\psi}_i(\mathbf{y}_i^1, \phi)$ is given by the inner summation in (3.12), with $\hat{\phi}$ replaced by ϕ, and the limit $\tilde{\boldsymbol{\Psi}}$ function is

$$\tilde{\psi}(\mathbf{y}_1, \phi) = \sum_{j=1}^{\infty} \phi^{j-1}\eta(y_1 - \phi y_0, y_{1-j} - \phi y_{-j}). \tag{3.13}$$

EXAMPLE 3. *GM estimates for the MA*(1) *model.* In order to motivate the definition of GM estimate for the MA(1) model, we first note that the least-squares estimate $\hat{\theta}_{\mathrm{LS}}$ of θ is a solution of the equation

$$\sum_{i=2}^{n} s_{i-1}^{*}(\hat{\theta}_{\mathrm{LS}}) r_i^{*}(\hat{\theta}_{\mathrm{LS}}) = 0, \tag{3.14}$$

where

$$s_i^{*}(\theta) = \sum_{j=0}^{i-1} (j+1)\theta^j y_{i-j} \tag{3.15}$$

and

$$r_i^{*}(\theta) = \sum_{j=0}^{i-1} \theta^j y_{i-j}. \tag{3.16}$$

It is easy to verify that when $y_i \equiv x_i$ with x_i the MA(1) model (3.5), $\lim_{i\to\infty} \operatorname{var} r_i^{*}(\theta) = 1$ and $\lim_{i\to\infty} \operatorname{var} s_i^{*}(\theta) = 1/(1-\theta^2)$. Therefore, by analogy with the AR(1) model, we define the GM estimate $\hat{\theta}$ of θ by

$$\sum_{i=2}^{n} \eta\left(r_i^{*}(\hat{\theta}), s_{i-1}^{*}(\hat{\theta})(1-\hat{\theta}^2)^{1/2}\right) = 0. \tag{3.17}$$

Here $\tilde{\psi}_i(\mathbf{y}_i^1, \theta)$ is given by the ith summand above, with $s_{i-1}^{*}(\theta)$ and $r_i^{*}(\theta)$ expressed in terms of the y_j for $1 \le j \le i$. The limit $\tilde{\Psi}$ function is

$$\tilde{\psi}(\mathbf{y}_1, \theta) = \eta\left(r_1(\theta), s_0(\theta)(1-\theta^2)^{1/2}\right), \tag{3.18}$$

where

$$s_i(\theta) = \sum_{j=0}^{\infty} (j+1)\theta^j y_{i-j}, \qquad r_i(\theta) = \sum_{j=0}^{\infty} \theta^j y_{i-j}. \tag{3.19}$$

EXAMPLE 4. *RA estimates for the MA*(1) *model.* The RA estimates $\hat{\theta}$ for the MA(1) model have exactly the same form of estimating equations as in the AR(1) case:

$$\sum_{i=3}^{n} \sum_{l=1}^{i-2} \hat{\theta}^{l-1} \eta\left(r_i(\hat{\theta}), r_{i-l}(\hat{\theta})\right) = 0 \tag{3.20}$$

with the residuals given by

$$r_i(\theta) = \sum_{j=0}^{i-1} \theta^j y_{i-j}. \tag{3.21}$$

It can be shown that when $\eta(\xi_1, \xi_2) = \xi_1 \xi_2$, $\hat{\theta}$ is asymptotically equivalent to the least-squares estimate.

The function $\tilde{\psi}_i(\mathbf{y}_i^1, \theta)$ is given by the inner summation of the estimating equation (3.20), with $\hat{\theta}$ replaced by θ. The limit $\tilde{\Psi}$ function is

$$\tilde{\psi}(\mathbf{y}_1, \theta) = \sum_{j=1}^{\infty} \theta^{j-1} \eta\left(\sum_{k=0}^{\infty} \theta^k y_{1-k}, \sum_{k=0}^{\infty} \theta^k y_{1-k-j} \right). \tag{3.22}$$

4. Influence functionals for time series.

4.1. **ICH** *for time series and the need for a new definition.* Since we deal only with estimates which are defined asymptotically by functionals $\mathbf{T} = \mathbf{T}(\mu)$, it might at first blush be tempting to simply apply Hampel's definition (1.1) in the time-series parameter-estimation setting. To do so, one would replace the univariate contaminating distribution F_γ by the process contamination measure $\mu_\gamma = (1-\gamma)\mu + \gamma\delta_{\mathbf{y}}$, where in general $\mathbf{y} = (\dots, y_1, y_0, y_{-1}, \dots) \in R^{-\infty,\infty}$, $\delta_{\mathbf{y}}$ is the unit mass at $\mathbf{y}$, and μ is a measure in $\mathbf{P}_{se}$. We assume that a $\tilde{\Psi}$ estimate $\mathbf{T} = \mathbf{T}(\mu)$ is defined by (3.3) for not only stationary and ergodic measures μ_x for the core process x_t in (2.2), but also for the contamination measures $\mu_\gamma = (1-\gamma)\mu_x + \gamma\delta_{\mathbf{y}}$, $0 \le \gamma \le 1$. Since $\delta_{\mathbf{y}} \notin \mathbf{P}_{se}$, this places some restriction on $\mathbf{y}$.

DEFINITION 4.1 (**ICH** *for time-series $\tilde{\Psi}$ estimates*). We define

$$\mathbf{ICH}(\mathbf{y}, \mathbf{T}, \mu) = \lim_{\gamma \to 0} \frac{\mathbf{T}(\mu_\gamma) - \mathbf{T}(\mu)}{\gamma}, \tag{4.1}$$

providing the limit exists.

Under suitable regularity conditions (4.1) and (3.3) yield

$$\mathbf{ICH}(\mathbf{y}_1) = \mathbf{ICH}(\mathbf{y}_1, \mathbf{T}, \mu_x) = -\mathbf{C}^{-1}\tilde{\Psi}(\mathbf{y}_1; \mathbf{t}_0), \tag{4.2}$$

where $\mathbf{t}_0 = \mathbf{T}(\mu_x)$ and the nonsingular matrix $\mathbf{C}$ is given by

$$\mathbf{C} = \left[(\partial/\partial \mathbf{t}) E\tilde{\Psi}(\mathbf{x}_1, \mathbf{t})\right]_{\mathbf{t}=\mathbf{t}_0}. \tag{4.2$'$}$$

This possibility is most tempting when $\mathbf{T}$ depends only on a finite-dimensional marginal measure μ^k, as in the case of GM/BIF estimates for autoregression, where the analogy with ordinary regression is suggestive. For such cases one would set $\mu_\gamma = \mu_\gamma^k = (1-\gamma)\mu_x^k + \gamma\delta_{\mathbf{y}_1^{-k+1}}$, where $\delta_{\mathbf{y}_1^{-k+1}}$ has all its mass at $\mathbf{y}_1^{-k+1}$. Such a definition was suggested by Martin and Jong (1977), Portnoy (1977), and Martin (1980), and pursued in a more serious vein by Künsch (1984), who focused on Hampel-type optimality results based on **ICH**. Künsch in fact proved that (4.2)–(4.2$'$) holds for pth-order autoregressions, and in that context also provided an empirical interpretation in the context of adding a single observation at the end of the series.

The **ICH** (4.1)–(4.2) does typically give the correct asymptotic variance–covariance matrix for $\mathbf{T}_n$:

$$\mathbf{V} = \mathbf{V}_o + \sum_{l=1}^{\infty} (\mathbf{V}_l + \mathbf{V}_l^{\mathrm{T}}), \tag{4.3}$$

where

$$\mathbf{V}_l = \operatorname{var}\left[\mathbf{ICH}(\mathbf{y}_1), \mathbf{ICH}(\mathbf{y}_{1+l})\right]. \tag{4.4}$$

In the case of estimating location with an ordinary M estimate $\mathbf{V}$ reduces to the expression obtained by Portnoy (1977). In the case of autoregression GM/BIF estimates, $\mathbf{V}$ coincides with the expression stated in Martin (1980) for the Mallows variant, and established rigorously by Bustos (1982) for a general class which includes both Mallows- and HKW-type estimates.

Unfortunately there is a very basic sense in which **ICH** is not the most appropriate definition for the time-series context: the definition does not correspond to any interesting contaminated time series. The reason is simple enough. One computes **ICH** by letting $\mu_\gamma = (1-\gamma)\mu + \gamma\delta_{\mathbf{y}}$, where $\delta_{\mathbf{y}}$ on $(R^{-\infty,\infty}, \mathscr{B})$ puts all its mass at the point $\mathbf{y} \in R^{-\infty,\infty}$. But the mixture structure of μ_γ implies that each sample path of the series is generated either by μ or by $\delta_{\mathbf{y}}$, and this hardly reflects the nature of any real contaminated series arising in practice.

4.2. *The time-series influence functional and its properties.*

DEFINITION 4.2 (*Time-series influence functional*). Suppose the estimator sequence $\{\mathbf{T}_n\}$ is specified asymptotically by a functional $\mathbf{T} = \mathbf{T}(\mu)$ defined on a subset $\mathbf{P}_o$ of $\mathbf{P}_{se}$, and suppose that μ_y^γ is given by (2.1)–(2.2). Then the *influence functional* **IF** of $\mathbf{T}$ is defined as

$$\mathbf{IF}\left(\mu_w, \mathbf{T}, \{\mu_y^\gamma\}\right) = \lim_{\gamma \to 0} \frac{\mathbf{T}(\mu_y^\gamma) - \mathbf{T}(\mu_y^o)}{\gamma} \tag{4.5}$$

provided the limit exists.

Note that the influence functional depends not only upon the estimator $\mathbf{T}$, the nominal model μ_x, and the contamination process measure μ_w as "main" argument, but also upon the particular *trajectory* or *arc* of contamination measures $\{\mu_y^\gamma\} = \{\mu_y^\gamma\colon 0 \le \gamma < 1\}$, as the fraction of contamination $g(\gamma) = \gamma + o(\gamma)$ tends to zero. By way of contrast the **ICH** for time series depends only on $\mathbf{T}$, the nominal measure μ, with $\mu = \mu_x$ in the present context, and the "main" argument $\delta_{\mathbf{y}}$. Correspondingly, **ICH** is a directional or Gateaux derivative with direction specified by μ_x and $\delta_{\mathbf{y}}$. In order to capture the essential features of time-series contamination in an *influence functional*, we take derivatives along particular arcs $\{\mu_y^\gamma\}$ to μ_x in $\mathbf{P}_o$.

It turns out that there is a rather simple connection between **IF** and **ICH** under certain conditions, and it is a connection which facilitates the computation of **IF**. The first theorem to follow gives sufficient conditions to insure that a trivial expansion argument will yield the key relation (4.6) below.

THEOREM 4.1. *Assume that* $\mathbf{T}$ *is a* $\tilde{\boldsymbol{\Psi}}$ *estimate with*

(a) $\lim_{\gamma \to 0} \mathbf{T}(\mu_{\mathrm{y}}^\gamma) = \mathbf{T}(\mu_{\mathrm{x}}) = \mathbf{t}_0$.

(b) *Put*

$$\mathbf{m}(\gamma, \mathbf{t}) = E\tilde{\boldsymbol{\Psi}}(\mathbf{y}_1^\gamma, \mathbf{t}).$$

There exists an $\varepsilon > 0$ *such that*

$$\mathbf{D}(\gamma, \mathbf{t}) = (\partial/\partial \mathbf{t})\mathbf{m}(\gamma, \mathbf{t})$$

exists for $0 \le \gamma \le \varepsilon$, $|\mathbf{t} - \mathbf{t}_0| < \varepsilon$, *and* $\mathbf{D}(\gamma, \mathbf{t})$ *is continuous at* $(0, \mathbf{t}_0)$. *Also* $\mathbf{C} = \mathbf{D}(0, \mathbf{t}_0)$ *is nonsingular.*

(c) $\lim_{\gamma \to 0} \mathbf{m}(\gamma, \mathbf{t}_0)/\gamma$ *exists.*

Then

$$\mathbf{IF}\left(\mu_w, \mathbf{T}, \{\mu_y^\gamma\}\right) = \lim_{\gamma \to 0} \frac{E\,\mathbf{ICH}(\mathbf{y}_1^\gamma)}{\gamma}. \tag{4.6}$$

COMMENT 4.1. For the case where $\tilde{\mathbf{\Psi}}$ depends on finitely many arguments the above expression coincides with (1.18) of Künsch (1984), which in that case is valid for more general contamination than those considered here.

Suppose that a $\tilde{\mathbf{\Psi}}$ estimate is selected with a view toward use in the i.i.d. setting. For example, $\tilde{\mathbf{\Psi}}$ might define an ordinary M estimate of location or scale. Then the ith summand in (3.1) depends only on the ith observation and $\mathbf{T}_n$ is permutation invariant by virtue of the equality

$$\tilde{\mathbf{\Psi}}_i(\mathbf{y}_i^1, \mathbf{t}) = \tilde{\mathbf{\Psi}}(y_i, \mathbf{t}), \qquad i = 1, 2, \dots, n.$$

The following corollary shows that **IF** is a strict generalization of **ICH** in the sense that the two coincide for such estimates when y_i^γ is an i.i.d. *pure replacement* process.

COROLLARY 4.1. *Suppose that the* $\tilde{\mathbf{\Psi}}$ *estimate satisfies*

$$\tilde{\mathbf{\Psi}}_i(\mathbf{y}_i^1, \mathbf{t}) = \tilde{\mathbf{\Psi}}(y_i, \mathbf{t}), \qquad \forall \mathbf{t},\ i = 1, 2, \dots, n,$$

and that y_i^γ *is given by* (2.2) *with* z_i^γ *i.i.d,* $g(\gamma) = \gamma$, *and* $w_i \equiv \xi$. *Then*

$$\mathbf{IF}\left(\mu_w, \mathbf{T}, \{\mu_y^\gamma\}\right) = \mathbf{IF}\left(\xi, \mathbf{T}, \{\mu_y^\gamma\}\right) = \mathbf{ICH}(\xi). \tag{4.7}$$

Assumption (b) in Theorem 4.1 is quite restrictive. The next theorem gives a useful set of conditions to insure that the relation (4.6) between **IF** and **ICH** holds for many types of estimates of interest, including GM and RA estimates, when the process $\{z_i^\gamma\}$ is independent of $\{x_i, w_i\}$ and has distribution $\mu_z^{k,\gamma}$ corresponding to patches of length k generated according to (2.4) of Section 2.2. The theorem also shows how to compute **IF** in these cases.

THEOREM 4.2. *Let* $\mathbf{T}$ *be a* $\tilde{\mathbf{\Psi}}$ *estimate with* $\mathbf{t}_0 = \mathbf{T}(\mu_x)$ *and suppose that* $\mu_{xwz} = \mu_{xw}\mu_z^{k,\gamma}$. *Assume* **ICH** *is given by* (4.2)–(4.2′) *and that*:

(a) $\mathbf{T}(\mu_y^\gamma) - \mathbf{t}_0 = O(\gamma)$.

(b) $E\tilde{\mathbf{\Psi}}(\mathbf{x}_1, \mathbf{t})$ *is differentiable at* $\mathbf{t} = \mathbf{t}_0$ *and the derivative matrix* $\mathbf{C}$ *given by* (4.2′) *is nonsingular.*

(c) *For* $m \ge 1$ *put*

$$H_m(\mathbf{t}) = \sup\left|E\tilde{\mathbf{\Psi}}\left(\mathbf{y}_1^{2-m}, \mathbf{y}_{1-m}, \mathbf{t}\right) - E\tilde{\mathbf{\Psi}}\left(\mathbf{y}_1^{2-m}, \mathbf{y}_{1-m}^*, \mathbf{t}\right)\right|,$$

where the supremum is with respect to every $\mathbf{y}_1^{2-m} = (y_1, y_0, \dots, y_{2-m})$, $\mathbf{y}_{1-m} = (y_{1-m}, y_{-m}, \dots)$, $\mathbf{y}_{1-m}^* = (y_{1-m}^*, y_{-m}^*, \dots)$ *such that each* y_{1-j} *and* y_{1-j}^* *may be* *either* x_{1-j} *or* w_{1-j}, *and put*

$$H_m^*(\varepsilon) = \sup_{|\mathbf{t}-\mathbf{t}_0|\le\varepsilon} H_m(\mathbf{t}).$$

There exists $\varepsilon_0 > 0$ *such that*

$$\sum_{m=1}^{\infty} H_m^*(\varepsilon_0) < \infty.$$

(d) *Put*

$$H_0(\mathbf{t}) = \sup E\tilde{\Psi}(\mathbf{y}_1, \mathbf{t}),$$

where the supremum is with respect to every $\mathbf{y}_1 = (y_1, y_0, \dots)$ *such that* y_{1-j} *may be either* x_{1-j} *or* w_{1-j}, *and put*

$$H_0^*(\varepsilon) = \sup_{|\mathbf{t}-\mathbf{t}_0|\le\varepsilon} H_0(\mathbf{t}).$$

There exists $\varepsilon_0 > 0$ *such that* $H_0^*(\varepsilon_0) < \infty$.

(e) *For any* $\mathbf{y}_1 = (y_1, y_0, \dots)$, *where each* $y_{1-\iota}$ *is either* $x_{1-\iota}$ *or* $w_{1-\iota}$, *we have*

$$\lim_{\mathbf{t}\to\mathbf{t}_0} \tilde{\Psi}(\mathbf{y}_1, \mathbf{t}) = \tilde{\Psi}(\mathbf{y}_1, \mathbf{t}_0) \quad a.s.$$

and there exists $\varepsilon > 0$ *such that*

$$E \sup_{|\mathbf{t}-\mathbf{t}_0|\le\varepsilon} |\tilde{\Psi}(\mathbf{y}_1, \mathbf{t})| < \infty.$$

Then (4.6) *holds, and*

$$\mathbf{IF}\left(\mu_w, \mathbf{T}, \left\{\mu_y^{k,\gamma}\right\}\right) = -\frac{1}{k}\mathbf{C}^{-1}\sum_{j=0}^{\infty}\mathbf{G}_j^k, \tag{4.8}$$

where

$$\mathbf{G}_j^k = \begin{cases} E\tilde{\Psi}\left(\mathbf{w}_1^{1-j}, \mathbf{x}_{-j}, \mathbf{t}_0\right) & \text{if } 0 \le j \le k-1, \\ E\tilde{\Psi}\left(\mathbf{x}_1^{1-j+k}, \mathbf{w}_{-j+k}^{1-j}, \mathbf{x}_{-j}, \mathbf{t}_0\right) & \text{if } j \ge k. \end{cases} \tag{4.8$'$}$$

COMMENT 4.2. We will see in Section 5 that the assumptions of this theorem are satisfied under general regularity conditions on the η function for GM and RA estimates for the AR(1) and MA(1) models (Theorems 5.2 and 5.4, respectively).

As in the case of the Hampel influence curve, boundedness of **IF** is of interest in connection with robustness. The following theorem gives sufficient conditions for boundedness of **IF**. We introduce some notation. Given an $n \times m$ matrix $\mathbf{A}$, let $\|\mathbf{A}\|$ be defined as $\sup\{|\mathbf{A}\mathbf{u}|: |\mathbf{u}| = 1, \mathbf{u} \in R^m\}$. Let N_0 be the set of nonnegative integers. Given a subset $I \subset N_0$, let $\mathbf{y}_{\iota, I} = \{y_{\iota-k}: k \in I\}$.

THEOREM 4.3. *Let* $\mathbf{T}$ *be a* $\tilde{\boldsymbol{\Psi}}$ *estimate such that* (4.6) *holds. Assume also*

(a) $$\tilde{\boldsymbol{\Psi}}(\mathbf{y}_1, \mathbf{t}) = \sum_{j=1}^{\infty} \boldsymbol{\pi}_j(\mathbf{y}_{1, I_j}, \mathbf{t}),$$

where the I_j *are subsets of* N_0 *such that the number of elements in each* I_j *is uniformly bounded by a finite integer* h, *and*

(b) $$E\boldsymbol{\pi}_j(\mathbf{x}_{1, I_j}, \mathbf{t}_0) = 0, \qquad \forall j.$$

$$\sup_{\mathbf{y}_{1, I_j}} |\boldsymbol{\pi}_j(\mathbf{y}_{1, I_j}, \mathbf{t}_0)| \le K_j$$

with

$$\sum_{j=1}^{\infty} K_j = K < \infty.$$

(c) $E\tilde{\boldsymbol{\Psi}}(\mathbf{x}_1, \mathbf{t})$ *is differentiable at* $\mathbf{t} = \mathbf{t}_0$, *and the derivative matrix* $\mathbf{C}$ *is nonsingular.*

Then

(4.9) $$\left|\mathbf{IF}\left(\mu_w, \mathbf{T}, \{\mu_y^{\gamma}\}\right)\right| \le 2hK\|\mathbf{C}^{-1}\|.$$

COMMENT 4.3. The above theorem applies to GM and RA estimates of autoregressive models if $\tilde{\boldsymbol{\Psi}}$ is bounded—see (3.7) and (3.13). However, boundedness of $\tilde{\boldsymbol{\Psi}}$ is not in general sufficient for the boundedness of **IF**. In Section 5.2 we give an MA(1) model example where estimates with bounded $\tilde{\boldsymbol{\Psi}}$ have unbounded **IF**, the reason being that the $\boldsymbol{\pi}_j$ depend on an infinite number of coordinates for moving-average models.

5. Influence functionals for AR(1) and MA(1) models with additive outliers. The computation of time-series influence functionals will be carried out for both GM estimates, denoted T^{GM} (cf. Examples 1 and 3) and RA estimates, denoted T^{RA} (cf. Examples 2 and 4). Throughout this section and the remainder of the paper the only outlier model we deal with is the AO model as described following (2.3), and with Gaussian AR(1) and MA(1) x_i processes (3.4) and (3.5). We selected the AO model for our computations primarily because it has been used in previous studies (e.g., Denby and Martin, 1979; Bustos and Yohai, 1986). We hope later on to make computations for pure replacement models, higher-order AR and MA models, etc.

Corresponding to the special cases treated, it is convenient to replace the notation $\mathbf{IF} = \mathbf{IF}(\mu_w, \mathbf{T}, \{\mu_y^{\gamma}\})$ by $\mathrm{IF}_{\mathrm{AO}, k} = \mathrm{IF}_{\mathrm{AO}, k}(\mu_v, T, \lambda)$ where $T = T^{\mathrm{GM}}$ or $T = T^{\mathrm{RA}}$ and $\lambda = \phi$ or θ. Here, the subscript k indicates the *patch* length for patchy outliers, and with μ_x fixed, specification of μ_v is equivalent to specification of μ_w for the AO model. When $k = 1$ we have *independent* outliers. We also replace **ICH** by ICH.

5.1. *AR*(1) *models*. We first state results concerning the ICH, and asymptotic variance (in central limit theorem form) of T^{GM} and T^{RA} at the Gaussian model. The need for the asymptotic variance under the nominal Gaussian model arises when comparing the IF's of different estimates: their tuning constants will be adjusted to obtain matched asymptotic efficiencies at the nominal model. The following assumptions will be used to prove these results:

(A1) $\eta(\cdot,\cdot)$ is continuous and odd in each variable.
(A2) $|\eta(u_1, u_2)| \le K|u_1|^{k_1}|u_2|^{k_2}$, where k_1 and k_2 are either 0 or 1.

(A3)
$$\eta_i(u_1, u_2) = \frac{\partial \eta(u_1, u_2)}{\partial u_i}, \qquad i = 1, 2,$$

are continuous and

$$|\eta_1(u_1, u_2)| \le K|u_2|^{h_1}, \qquad |\eta_2(u_1, u_2)| \le K|u_1|^{h_2},$$

where h_1 and h_2 are either 0 or 1.

(A4)

(5.1)
$$B = E\{v \cdot (\partial/\partial x)\eta(x, v)|_{x=u}\} \ne 0,$$

where u and v are independent $N(0,1)$ random variables.

Observe that the LS estimates satisfy (A2) with $k_1 = k_2 = 1$ and (A3) with $h_1 = h_2 = 1$. M estimates with bounded ψ and ψ' satisfy (A1) with $k_1 = 0$, $k_2 = 1$ and (A3) with $h_1 = 1$ and $h_2 = 0$. GM and RA estimates with bounded η satisfy (A2) with $k_1 = k_2 = 0$, and if they are of the Mallows types with ψ' bounded, they satisfy (A3) with $h_1 = h_2 = 0$, while if they are of the Hampel–Krasker type with Ψ' bounded, (A3) is satisfied with $h_1 = h_2 = 1$.

THEOREM 5.1.

(i) *Under* (A1), (A3), *and* (A4) *we have*

(5.2)
$$\text{ICH}(\mathbf{y}_1, T^{\text{GM}}, \phi) = \frac{(1-\phi^2)^{1/2}}{B}\eta\left(r_1(\phi), y_0(1-\phi^2)^{1/2}\right),$$

(5.3)
$$\text{ICH}(\mathbf{y}_1, T^{\text{RA}}, \phi) = \frac{1-\phi^2}{B}\sum_{j=1}^{\infty}\phi^{j-1}\eta\left(r_1(\phi), r_{1-j}(\phi)\right),$$

where $r_i(\phi) = y_i - \phi y_{i-1}$.

(ii) *Let* T_n *denote either* T_n^{GM} *or* T_n^{RA}, *and set*

(5.4)
$$A = E\eta^2(u, v)$$

with u, v *independent* $N(0,1)$ *random variables. If* (A1)–(A4) *and the Gaussian* $AR(1)$ *model* (3.4) *hold, then*

$$n^{1/2}(T_n - \phi) \to_L N(0, V)$$

with

(5.5)
$$V = (1-\phi^2)\frac{A}{B^2}.$$

THEOREM 5.2. *Suppose that the AO model holds, i.e.,* $w_t = x_t + v_t$ *in* (2.2), *with* v_t *independent of* x_t. *Assume* (A1)–(A4) *and that* $E|v_1|^{k_1+k_2} < \infty$, *where* k_1 *and* k_2 *are as in* (A2).

(i) *For independent outliers we have*

$$\text{(5.6)} \quad \mathrm{IF}_{\mathrm{AO},1}(u_v, T^{\mathrm{GM}}, \phi) = \frac{(1-\phi^2)^{1/2}}{B} E\eta\big(u_1 - \phi v_0, (x_0 + v_0)(1-\phi^2)^{1/2}\big)$$

and

$$\text{(5.7)} \quad \mathrm{IF}_{\mathrm{AO},1}(\mu_v, T^{\mathrm{RA}}, \phi) = \frac{1-\phi^2}{B} E\eta(u_1 - \phi v_0, u_0 + v_0).$$

(ii) *For patches of outliers of length* $k \geq 2$ *we have*

$$\text{(5.8)} \quad \begin{aligned} \mathrm{IF}_{\mathrm{AO},k}(\mu_v, T^{\mathrm{GM}}, \phi) = {} & \frac{(1-\phi^2)^{1/2}}{kB} \\ & \times \Big[(k-1)E\eta\big(u_1 + v_1 - \phi v_0, (x_0 + v_0)(1-\phi^2)^{1/2}\big) \\ & \quad + E\eta\big(u_1 - \phi v_0, (x_0 + v_0)(1-\phi^2)^{1/2}\big)\Big] \end{aligned}$$

and

$$\text{(5.9)} \quad \begin{aligned} \mathrm{IF}_{\mathrm{AO},k}(\mu_v, T^{\mathrm{RA}}, \phi) = {} & \frac{1-\phi^2}{kB} \\ & \times \Bigg[\sum_{h=1}^{k-2} (k-h-1)\phi^{h-1} E\eta(u_1 + v_1 - \phi v_0, u_{1-h} + v_{1-h} - \phi v_{-h}) \\ & + \sum_{h=1}^{k-1} \phi^{h-1} E\eta(u_1 + v_1 - \phi v_0, u_{1-h} + v_{1-h}) \\ & + \sum_{h=1}^{k-1} \phi^{h-1} E\eta(u_1 - \phi v_0, u_{1-h} + v_{1-h} - \phi v_{-h}) \\ & + \phi^{k-1} E\eta(u_1 - \phi v_0, u_{1-k} + v_{1-k})\Bigg]. \end{aligned}$$

COMMENT 5.1. The expectations in (5.6)–(5.9) are with respect to the measure $\mu_{xv} = \mu_x \mu_v$ where μ_x yields all the necessary joint distributions for the x_t and u_j in the AR(1) model. Here the measure $\mu_v \in \mathbf{P}_{se}$ is quite general. We specialize to the leading case of degenerate measures δ_ζ, corresponding to $v_t \equiv \zeta$, when computing IF's in Section 5.3.

5.2. *MA*(1) *models*. The following theorem gives the ICH's and asymptotic variances of T^{RA} and T^{GM} for the MA(1) Gaussian model, as given in Examples 3 and 4, respectively, of Section 3. The scalar-valued limit $\tilde{\psi}$ functions for T^{RA} and T^{GM} are given by (3.18) and (3.22), respectively.

THEOREM 5.3. (i) *Under* (A1), (A3), *and* (A4) *we have*

$$\text{ICH}(\mathbf{y}_1, T^{\text{GM}}, \theta) = -\frac{(1-\theta^2)^{1/2}}{B}\eta\big(r_1(\theta), s_0(\theta)(1-\theta^2)^{1/2}\big), \tag{5.10}$$

$$\text{ICH}(\mathbf{y}_1, T^{\text{RA}}, \theta) = -\frac{(1-\theta^2)}{B}\sum_{j=1}^{\infty}\theta^{j-1}\eta\big(r_1(\theta), r_{1-j}(\theta)\big), \tag{5.11}$$

where B is given in (A4), $r_\iota(\theta)$ *and* $s_\iota(\theta)$ *are given by* (3.19).

(ii) *If* T_n *represents the GM or RA estimate, and* (A1)–(A4) *and the Gaussian MA*(1) *model* (3.5) *hold, then*

$$n^{1/2}(T_n - \theta) \to_d N\big(0, (1-\theta^2)A/B^2\big), \tag{5.12}$$

where A is given by (5.4), *and B is given in* (A4).

The following theorem gives the IF's of T^{GM} and T^{RA} for patches of length one (i.e., independent outliers).

THEOREM 5.4. *Assume* (A1)–(A4), *and that the AO model holds, i.e.,* $w_i = x_\iota + v_\iota$, *with* $v_\iota \sim \mu_v$ *and independent of* x_ι. *Further suppose that the process* $\{z_\iota^\gamma\}$ *is independent of the processes* $\{x_\iota, v_\iota\}$, *with* z_ι^γ *an i.i.d. Bernoulli sequence. Assume also that* $E|v_1|^{h+1} < \infty$, *where* $h = \max(h_1, h_2)$ *with* h_1, h_2 *as in* (A3). *Then*

$$\begin{aligned}\text{IF}_{\text{AO},1}(\mu_v, T^{\text{GM}}, \theta) &= -\frac{(1-\theta^2)^{1/2}}{B}\\ &\quad\times\sum_{j=1}^{\infty}E\eta\big(u_1 + \theta^j v_1, u_2 + j(1-\theta^2)^{1/2}\theta^{j-1}v_1\big)\end{aligned} \tag{5.13}$$

and

$$\begin{aligned}\text{IF}_{\text{AO},1}(\mu_v, T^{\text{RA}}, \theta) &= -\frac{(1-\theta^2)}{B}\\ &\quad\times\sum_{j=1}^{\infty}\theta^{j-1}\sum_{\iota=j}^{\infty}E\eta\big(u_1 + \theta^i v_1, u_2 + \theta^{i-j}v_1\big),\end{aligned} \tag{5.14}$$

where u_1 *and* u_2 *are independent* $N(0,1)$ *random variables, with* u_1, u_2 *independent of* v_1.

COMMENT 5.2. From formulas (5.13) and (5.14) it is easy to see that for the MA(1) model the influence functional of GM and RA estimates is unbounded when η is monotone but *bounded*. Just take the supremum of the above influence curves over ζ, with $\mu_v = \delta_\zeta$. Thus boundedness of the $\tilde{\psi}$ function in (3.18) and (3.22) does not insure boundedness of the IF for MA(1) models. This is a general feature of models with moving-average components.

However, it is possible to show that when η is of the Mallows type (3.8) or HKW type (3.9) based on redescending ψ, e.g., ψ_{BS} given by (5.15), the corresponding RA and GM estimates have bounded IF. These results are illustrated in calculations to follow.

5.3. *Some influence curve computations.* Although general measures μ_v are used in the IF expressions of the preceding subsection, a leading case for expressing the intuitive notion of the "influence" of a configuration of contamination points is obtained by using a degenerate measure for μ_v (cf. Sections 2.1 and 2.3 on configurations). Thus for both i.i.d. and patch "configurations" we shall let $P(v_i = \zeta) = 1$, for a constant ζ (among other possibilities one might let $P(v_i = \zeta) = P(v_i = -\zeta) = \frac{1}{2}$). This allows one to step down from the abstract view of the IF as a functional on measure space. One can now compute and plot IF's as a function of the contamination value ζ, thus retaining the rich heuristics attraction of the ICH. Correspondingly, we use the term *influence curve*, IC = IC(ξ), to describe this special case of an influence functional

We calculated IC(ζ) for least-squares (LS), GM, and RA estimates of the HKW type at the following AO models: (i) AR(1) with *both* independent and patch outliers, and (ii) MA(1) with *only* independent outliers. Two psi functions are used for each of the choices of η, namely the Tukey redescending bisquare function

$$\psi_{BS,a}(u) = \begin{cases} u\left(1 - (u/a)^2\right)^2, & |u| \le a, \\ 0, & |u| > a, \end{cases} \tag{5.15}$$

and the Huber function

$$\psi_{H,a}(u) = \min(a, \max(u, -a)). \tag{5.16}$$

The tuning constants a were adjusted for each estimate to obtain 95% efficiency at a perfectly observed Gaussian AR(1) process. The values of the constants are given in Table 5.1.

The results of these IC calculations are shown in Figures 1–3. Figure 1 shows the AR(1) results, with $\phi = 0.5$, for independent outliers and patches of length 20. Figure 2 is the same except that $\phi = 0.9$. The least-squares influence curve is quadratically unbounded in both cases. The general messages for the robust estimates are clear: (i) the bisquare psi function is preferred to the Huber psi function; (ii) the RA estimates are preferred over the GM estimates for independent outliers, while the reverse is true for long patches.

TABLE 5.1
Tuning constants

	HKW estimate	**Mallows estimate**
Ψ_H	2.52	1.65
Ψ_{BS}	9.36	5.58

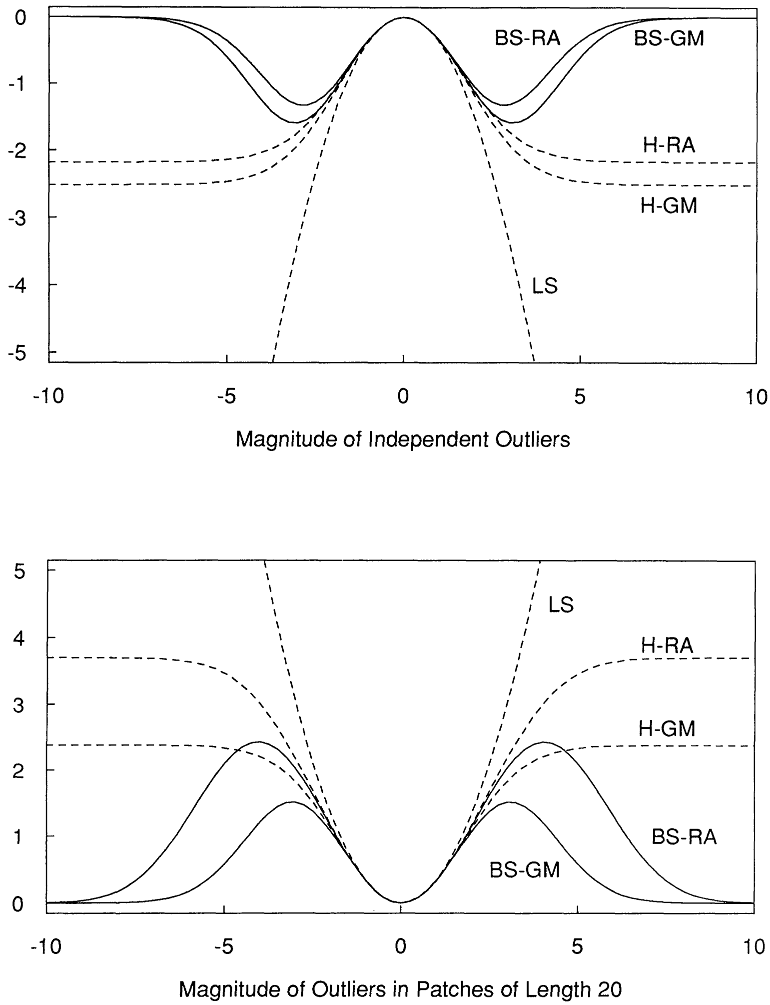

FIG. 1. *Influence curves for the AR*(1) *model. Hampel–Krasker–Welsch estimates at* $\phi = 0.5$.

Our preference orderings here are based on the gross-error sensitivities (GES's) of the IC's, the GES's for these particular examples being simply the supremum of $|\mathrm{IC}(\zeta)|$. The GES's here have property (P3), just as in the case of the GES for Hampel's influence curve ICH.

Figure 3 shows IC's for the AO MA(1) model with $\theta = -0.5$ and $\theta = -0.9$ (with our sign convention this gives positive correlation for the x_t process at lag-one). The results are in keeping with Comment 5.2: GM and RA estimates based on the monotone Huber psi functions have *unbounded* IC's (though apparently not quadratically unbounded as in the case of LS), while the bisquare ψ function leads to bounded IC's for the MA(1) model. Also, the GM estimate seems to be preferable to the RA estimate.

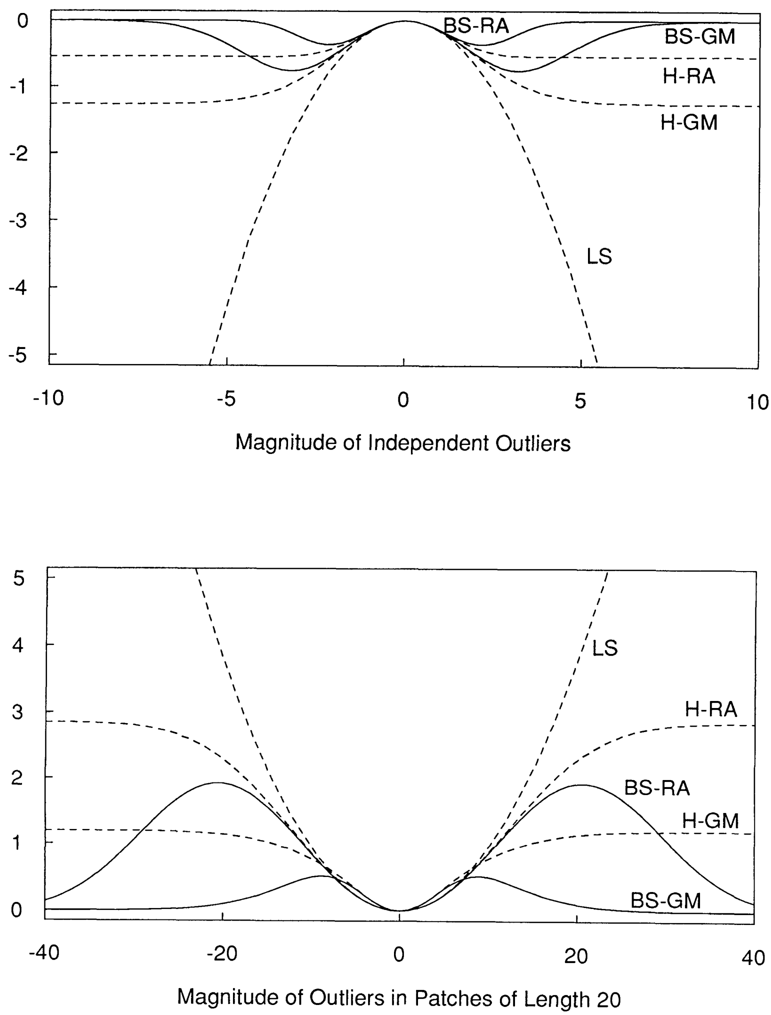

FIG. 2. *Influence curves for the $AR(1)$ model. Hampel–Krasker–Welsch estimates at* $\phi = 0.9$.

This last observation surprised us, as we had been somewhat pessimistic about using GM estimates for MA and ARMA models since such estimates seemed particularly natural only for AR models. We are now motivated to take the possibility of using GM estimates for ARMA models more seriously, and undertake a careful study.

We have carried out a parallel set of IC calculations based on GM and RA estimates of the Mallows type. The Mallows type IC's are displayed in Figures 4–6 of Martin and Yohai (1984a), and differ from the HKW type IC's presented here only by virtue of having slightly different shapes and slightly larger GES's.

COMMENT 5.3. Some proposed robust ARMA model parameter estimates are not tractable with regard to obtaining closed-form expressions for their influence

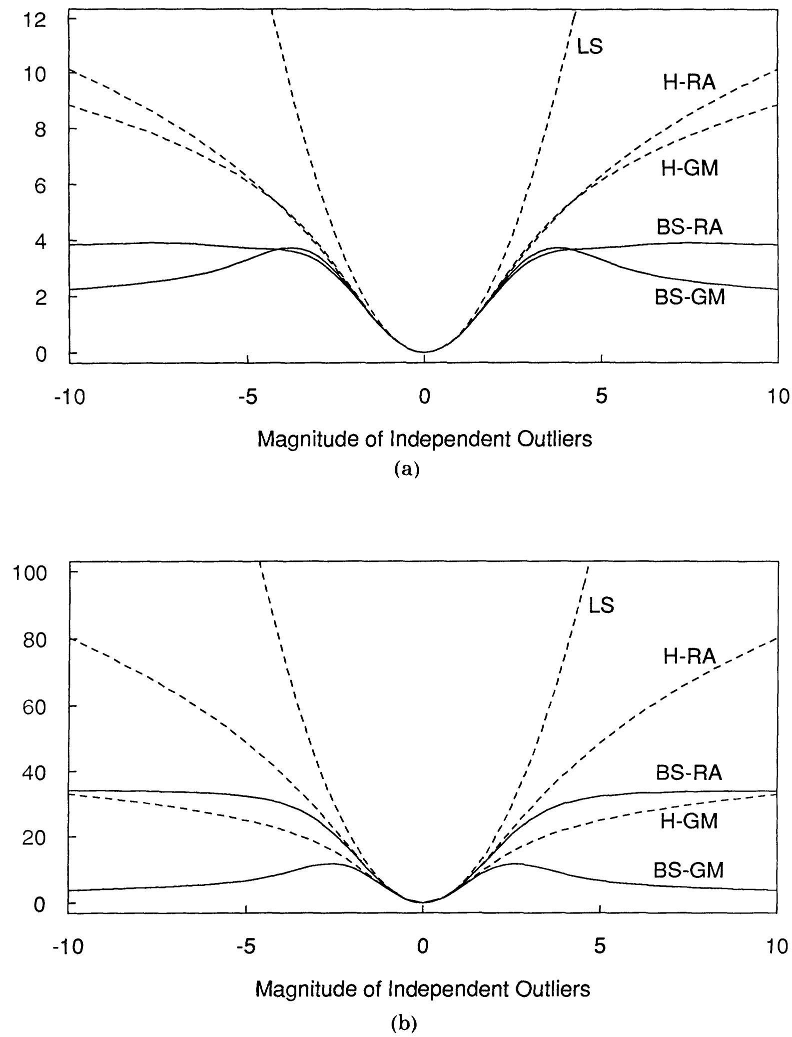

FIG. 3. *Influence curves for the MA*(1) *model. Hampel-Krasker–Welsch estimates* (a) *at* $\theta = -0.5$ *and* (b) $\theta = -0.9$.

functionals or influence curves. This is the case for example with the AM estimates based on robust filter cleaners (see, for example, Kleiner, Martin, and Thomson, 1979; Martin, 1981; Martin and Yohai, 1985). However, one can estimate influence functionals for such estimates via simulation. Some work has been completed along these lines, and will be reported elsewhere.

6. Gross-error sensitivity. In this section we give a general definition of gross-error sensitivity (**GES**) based on the **IF**. Specific results are then given for the **GES**'s of GM and RA estimates of the first-order autoregression parameter.

6.1. *The gross-error sensitivity.* Suppose that the contamination process y_i^γ is given by (2.2) and its related assumptions. Then for a fixed set of measures $\{\mu_{xwz}^\gamma\} = \{\mu_{xwz}^\gamma: 0 \le \gamma < 1\}$, we have a particular arc $\{\mu_y^\gamma\} = \{\mu_y^\gamma: 0 \le \gamma < 1\}$ of contaminated process measures in $\mathbf{P}_{se}$. Suppose that the **IF** (4.5) exists for a given *family* **P** of arcs $\{\mu_y^\gamma\}$ with each arc in $\mathbf{P}_{se}$. This family **P** will be generated by letting the contamination process measure μ_w vary over a family $\mathbf{P}_w$, in a manner consistent with the dependency structure $\{\mu_{xwz}^\gamma\}$, while $\{\mu_z^\gamma\} = \{\mu_z^\gamma: 0 \le \gamma < 1\}$ and μ_x are fixed. Often $\{\mu_z^\gamma\}$ will be independent of μ_{xw}, and then **P** will be generated by letting μ_w vary over $\mathbf{P}_w$ in a manner consistent with the dependence structure of μ_{xw}. For the pure replacement model given in Section 2.2, $\mu_{xw} = \mu_x \mu_w$ so **P** is then determined by simply letting μ_w range over $\mathbf{P}_w$. In the AO model where $w_i = x_i + v_i$ with x_i and v_i independent, the measures $\mu_{xw} = \mu_{x,x+v}$ are specified by μ_x and μ_v, and **P** is generated by letting μ_v range over a prescribed family $\mathbf{P}_v$.

Definition 6.1. The gross-error sensitivity (**GES**) of an estimate **T** at the family **P** of arcs $\{\mu_y^\gamma\}$ is

$$\mathbf{GES}(\mathbf{P}, \mathbf{T}) = \sup_{\{\mu_y^\gamma\} \in \mathbf{P}} \left| \mathbf{IF}\left(\mu_w, \mathbf{P}, \{\mu_y^\gamma\}\right) \right|. \tag{6.1}$$

Comment 6.1. Since **ICH** has leading argument **y** by virtue of being a directional derivative determined by the point mass contamination measure $\delta_{\mathbf{y}}$ Hampel's (1974) **GESH** is the supremum over all **y** of $|\mathbf{ICH}(\mathbf{y})|$. Our **IF** depends on the arc $\{\mu_y^\gamma\}$, with each $\mu_w \in \mathbf{P}_w$ specifying a particular $\{\mu_y^\gamma\} \in \mathbf{P}$, and so our **GES** involves the supremum over arcs $\{\mu_y^\gamma\}$ in **P**.

Comment 6.2. In either a pure replacement model or an AO model where the family of arcs **P** is generated by $\mathbf{P}_w$ or $\mathbf{P}_v$, respectively, a leading case of **GES** is obtained when $w_i \equiv v_i \equiv \zeta$, and correspondingly $\mathbf{P}_w = \mathbf{P}_v = \{\delta_\zeta\}$ where δ_ζ is the point mass on R^∞. In this case we replace **IF** by **IC** (for *influence curve*), replace the argument μ_w by ζ, take the supremum over all ζ, and replace the **GES** argument **P** by $\{\mu_z^\gamma \mu_x\}$.

6.2. *GES computations.* In Table 6.1 below we give GES's corresponding to the Mallows estimate IC's computed for AR(1) models in Section 5.3. Specific formulas which allow one to compute these GES's are derived in Martin and Yohai (1984b). The estimates are matched by the choice of tuning constants a to have the same asymptotic efficiencies at the Gaussian model. This means that $\sup_r |\psi_{\mathrm{BS}, a_1}(r)| > \sup_r |\psi_{\mathrm{H}, a_2}(r)|$, and correspondingly, $\sup_{u,v} |\eta_{\mathrm{BS}, a_1}(u, v)| > \sup_{u,v} |\eta_{\mathrm{H}, a_2}(u, v)|$. Yet, as Table 6.1 shows, the GES's for the bisquare psi function are often smaller than those for the Huber psi function.

The GES's for the HKW estimates are smaller than those of the Mallows estimates, but the differences are only slight.

TABLE 6.1
AR(1) GES's for additive outliers

	$\phi = 0.5$		$\phi = 0.9$	
Estimator	$k = 1$	$k = 20$	$k = 1$	$k = 20$
LS	∞	∞	∞	∞
GM-H	−2.8	2.6	−1.4	1.3
RA-H	−2.5	4.3	−0.6	3.3
GM-BS	−1.9	1.6	−1.8	0.4
RA-BS	−1.5	3.1	−1.5	2.5

TABLE 6.2
MA(1) GES's for additive outliers

Estimator	$\theta = -0.5$	$\theta = -0.9$
LS	∞	∞
GM-H	∞	∞
RA-H	∞	∞
GM-BS	4.0	10.0
RA-BS	3.8	38.0

COMMENT 6.3. In Martin and Yohai (1984b) it is shown that for independently located additive outliers, RA estimates have smaller GES's than GM estimates. On the other hand, for long patches GM estimates are better. These properties are reflected in Table 6.1.

Table 6.2 gives GES's corresponding to the IC's computed for MA(1) models with $k = 1$ (independent outliers) in Section 5.3. The striking feature here was already evident in the IC calculations: For the MA(1) model, a redescending psi function ψ is needed to obtain a bounded IC. Although the GM-BS estimate is a bit worse than the RA estimate at $\theta = -0.5$, it is much better at $\theta = -0.9$ where the x_t process is correspondingly more highly correlated.

7. An optimality property of generalized RA estimates. As one application of the **IF** in constructing good estimates, we define here a class of generalized RA estimates (GRA estimates) for the AR(1) model (3.4) and show that a member of this class has a certain optimality property which we shall establish. A GRA estimate for the AR(1) model is a $\tilde{\mathbf{\Psi}}$ estimate with $\tilde{\psi}_t$ of the form

$$\tilde{\psi}_i^{\mathrm{GRA}}(\mathbf{y}_1, \phi) = \sum_{j=1}^{i-2} \eta_j\big(r_1(\phi), r_{1-j}(\phi), \phi\big), \qquad i \geq 3,$$

where $r_t(\phi) = y_t - \phi y_{t-1}$, and the limit $\tilde{\mathbf{\Psi}}$ function is assumed to be

$$\tilde{\psi}^{\mathrm{GRA}}(\mathbf{y}_1, \phi) = \sum_{j=1}^{\infty} \eta_j\big(r_1(\phi), r_{1-j}(\phi), \phi\big). \tag{7.1}$$

The AR(1) RA estimate of Example 3 in Section 2 has ψ function (3.13), which is a special case of the above GRA ψ function.

We will show that for a large subclass of general $\tilde{\Psi}$ estimators there exist GRA estimates with smaller asymptotic variance, and the same IF at AO models, when the outliers are independent and x_t is a Gaussian AR(1) process.

Consider a general $\tilde{\psi}$ estimate with limit $\tilde{\psi}$ function expressed in the form

$$\tilde{\psi}(\mathbf{y}_1, \phi) = \tilde{\psi}^+(r_1(\phi), r_0(\phi), \dots, \phi). \tag{7.2}$$

We will define our "optimal" GRA estimate $\tilde{\psi}^*$ by (7.1) with the particular η_j functions

$$\eta_j^*(u_1, u_{1-j}, \phi) = E\left[\tilde{\psi}^+(u_1, u_0, \dots, \phi) | u_1, u_{1-j}\right], \tag{7.3}$$

where the u_i are the autoregression innovations in (3.4). Call T the estimate based on $\tilde{\psi}$, and T^* the GRA estimate based on $\tilde{\psi}^*$.

We will use the following assumptions:

(C1) $\tilde{\psi}^+$ is differentiable in each variable, and for all $j \geq 1$, $k = 1, 1-j$,

$$(\partial/\partial u_k)\eta_j^*(u_1, u_{1-j}, \phi) = E\left[(\partial/\partial u_k)\tilde{\psi}^+(u_1, u_0, u_{-1}, \dots, \phi) | u_1, u_{1-j}\right],$$

and

$$(\partial/\partial t)\eta_j^*(u_1, u_{1-j}, t)\Big|_{t=\phi} = E\Big[(\partial/\partial u_k) \times \tilde{\psi}^+(u_1, u_0, u_{-1}, \dots, t)\Big|u_1, u_{1-j}\Big]\Big|_{t=\phi}.$$

(C2) $\tilde{\psi}^+(-a_1, a_2, a_3, \dots, \phi) = -\tilde{\psi}^+(a_1, a_2, a_3, \dots, \phi)$.

(C3) $\tilde{\psi}^+(a_1, -a_2, -a_3, \dots, \phi) = -\tilde{\psi}^+(a_1, a_2, a_3, \dots, \phi)$.

(C4) $\tilde{\psi}(\mathbf{y}_1, \phi)$ is bounded in $\mathbf{y}_1$ for each $\phi \in (-1, 1)$.

(C5) The order of integration and differentiation in (4.2′) may be interchanged, and with $\mathrm{D} = (\partial/\partial\phi)$,

$$C_{\tilde{\psi}} = E\mathrm{D}\tilde{\psi}(\mathbf{x}_1, \phi) \neq 0.$$

The same is true of $\tilde{\psi}^*$ in place of $\tilde{\psi}$.

(C6) T and T^* satisfy (4.6), and $\tilde{\psi}$ satisfies the conditions of Theorem 4.2.

(C7) T and T^* are asymptotically normal with variances given by (4.3)–(4.4).

Notice that the GM estimates of the Mallows and Hampel type with odd ψ functions satisfy (C2) and (C3). (C2) guarantees the Fisher consistency of the estimate when the u_i's have a symmetric distribution.

The following theorem gives the optimality property of the GRA estimate T^* based on $\tilde{\psi}^*$.

THEOREM 7.1. *Assume* (C1)–(C7), *and suppose that* x_t *is a Gaussian* $AR(1)$ *process. Then*

(i) $\mathrm{AVAR}(T^*) \leq \mathrm{AVAR}(T)$, *where* AVAR *denotes asymptotic variance when* $y_t \equiv x_t$.

(ii) *Suppose that* v_i *is independent of* x_ι. *Then*

$$\mathrm{IF}_{\mathrm{AO},1}(\mu_v, T^*, \phi) = \mathrm{IF}_{\mathrm{AO},1}(\mu_v, T, \phi).$$

EXAMPLE 7.1. For a GM estimate of the Mallows type we have

$$\tilde{\psi}(\mathbf{y}_1, \phi) = \psi(y_1 - \phi y_0)\psi\left(y_0(1-\phi^2)^{1/2}\right)$$

and the corresponding GRA estimate is given by

$$\eta_j(r_1, r_{1-j}, \phi) = \psi(r_1)\lambda\left(d_j(\phi)r_{1-j}, s_j(\phi)\right),$$

where

$$d_j(\phi) = \phi^{j-1}(1-\phi^2)^{1/2}, \qquad s_j(\phi) = \left[1-(1-\phi^2)\phi^{2(j-1)}\right]^{1/2}, \tag{7.4}$$

$$\lambda(a, b) = E\psi(a + bu),$$

with $u \sim N(0,1)$.

EXAMPLE 7.2. For the HKW-type GM estimate

$$\tilde{\psi}(\mathbf{y}_1, \phi) = \psi\left((y_1 - \phi y_0)y_0(1-\phi^2)^{1/2}\right)$$

and the corresponding GRA estimate is given by

$$\eta_j(r_1, r_{1-j}, \phi) = \lambda\left(d_j(\phi)r_{1-j} + s_j(\phi)r_1\right).$$

We note that if ψ is of the Huber type $\psi_{\mathrm{H},c}$, then

$$\lambda(a, b) = |b|\left[f_N\left(\frac{c+a}{|b|}\right) - f_N\left(\frac{c-a}{|b|}\right)\right] + (c+a)F_N\left(\frac{c+a}{|b|}\right)$$
$$-(c-a)F_N\left(\frac{c-a}{|b|}\right) - a,$$

where f_n and F_N are the standard normal density and distribution functions, respectively.

8. Generalizations and further applications. In defining the influence functional for time series, we have for convenience concentrated on: (i) the use of a *particular* type of contamination model, namely the general replacement model (2.2), and (ii) application of the influence functional to the study of some *particular* estimates of AR(1) and MA(1) parameter estimates. Neither of these two narrow points of focus adequately reflects the potential utility of the time-series influence functional as a tool for studying the effects of many types of contamination on statistics used in a wide variety of time-series problems.

With regard to contamination type, any realistic model of contamination can be used for which the derivative (4.5) defining the influence functional exists. It would of course be extremely helpful if the **IF** admits a tractable analytic form, but this is not absolutely essential (see comment at end of Section 5).

As for **IF**'s for other time-series statistics, the following two examples should give some indication of the range of possibilities which remain to be explored.

8.1. *Testing for white noise.* Given a zero mean time series $y_1, \dots, y_n$ with measure μ_y, one can test for whiteness of the series using the statistic

$$V_n^L = \sum_{l=1}^{L} r_l^2,$$

where

$$r_l = \frac{\sum_{t=l+1}^n y_t y_{t-l}}{\sum_{t=1}^n y_t^2}$$

is the lag-l autocorrelation estimate. The functional $T_{W,L} = T_{W,L}(\mu_y)$ associated with V_n^L is

$$T_{W,L} = \sum_{l=1}^{L} \rho_l^2,$$

where

$$\rho_l = \frac{Ey_1 y_{1-l}}{Ey_1^2}$$

is the lag-l autocorrelation estimate.

Under the null hypothesis that $\mu_y = \mu_y^0$ is a white-noise measure, nV_n^L converges in law to a chi-squared distribution χ_L^2, with L degrees of freedom. In this case it is easy to check that for the general replacement model (2.2)

$$\mathrm{IF}\left(\mu_w, T_{W,L}\{\mu_y^\gamma\}\right) = 0$$

for any μ_w and any arc $\{\mu_y^\gamma\}$ such that $\mu_y^\gamma \to \mu_y^0$ as $\gamma \to 0$. This is the usual von Mises expansion type of result for a statistic having an asymptotic chi-squared distribution. One therefore expects a nonvanishing second derivative, and hence we define the *second-order* influence functional for such cases as

$$\mathrm{IF}^{(2)}\left(\mu_w, T_{W,L}, \{\mu_y^\gamma\}\right) = \left(\partial^2/\partial\gamma^2\right)T\left(\mu_y^\gamma\right)\big|_{\gamma=0}.$$

For contaminated processes μ_y^γ with patches of length k, one finds that

$$\mathrm{IF}_k^{(2)}\left(\mu_w, T_{W,L}, \{\mu_y^\gamma\}\right) = \frac{1}{k^2\sigma_x^4}\sum_{j=1}^{L}\left[\min(k,j)\left(Ex_1w_{1-j} + Ew_1x_{1-j}\right) + \max(k-j,0)Ew_1w_{1-j}\right]^2.$$

When x_t is independent of w_{t-l}, $|l| = 1, 2, \dots,$ this expression reduces to

$$\mathrm{IF}_k^{(2)}\left(\mu_w, T_{W,L}, \{\mu_y^\gamma\}\right) = \frac{1}{k^2\sigma_x^4}\sum_{l=1}^{\min(L,k-1)}(k-l)^2\left(Ew_1w_{1-l}\right)^2.$$

In the AO case where $w_t = x_t + v_t$, with x_t white noise and v_t independent of x_t, we have

$$Ew_1w_{1-l} = Ev_1v_{1-l}, \qquad l \geq 1.$$

If $v_t \equiv \xi$, this gives the influence curves

$$\mathrm{IC}_k^{(2)}(\xi) = \frac{\xi^4}{\sigma_x^4}\frac{1}{k^2}\sum_{l=1}^{\min(L,k-1)}(k-l)^2, \qquad k = 1,2,\dots .$$

We have $\mathrm{IC}_1^{(2)}(\xi) \equiv 0$ as expected, and $\mathrm{IC}_k^{(2)}(\xi)$ increases with k until $k = L + 1$, as is intuitively reasonable.

Of course, since $\mathrm{IC}_k^{(2)}$ is unbounded, the test statistic V_n^L is not robust, and one is motivated to find a robust alternative.

COMMENT 8.1. Since the same test can be obtained using different test statistics which have different influence curves, it is necessary to standardize a test statistic in order to make fair comparisons in terms of influence curves. This can be done for example as described by Ronchetti (1982) and Lambert (1981). Similar standardizations should be considered in order to convert $\mathrm{IF}_k^{(2)}$ to an influence curve for tests.

8.2. *Spectral density estimates.* Let $S(F) = S(f, \mu_y)$ denote the spectral density functional of a stationary, zero mean process y_t with measure μ_y. Of course $S(f)$ in fact depends only upon second-order properties of y_t. It is common practice to estimate $S(f)$ with a smoothed version $S_n(f)$ of the periodogram based on "tapered" data, smoothing being needed to obtain consistency. Such estimates may be written as

$$S_n(f) = \sum_{l=-n}^{n} w_n(l)\hat{R}_n(l)e^{\iota 2\pi f l},$$

where $w_n(l)$ is an appropriate "lag window" and

$$\hat{R}_n(l) = \frac{1}{n}\sum_{t=|l|+1}^{n}(y_t - \bar{y})(y_{t-|l|} - \bar{y})$$

is an estimate of the lag-l covariance $R(l) = R(l, \mu_y)$ of the process y_t. Under certain conditions $\hat{R}_n(l)$ is a consistent estimate of $R(l)$, $l = 0, 1, \dots,$ and $S_n(f)$ is a consistent estimate of $S(f)$. Thus, the functional associated with $S_n(f)$ is

$$S(f) = \sum_{l=-\infty}^{\infty} R(l)e^{\iota 2\pi f l}.$$

The spectral density $S(f)$ is an infinite-dimensional parameter and we get an infinite-dimensional influence functional $\mathbf{IF}(\mu_w, S, \{\mu_y^\gamma\}) = \mathbf{IF}(\mu_w, f, S, \{\mu_y^\gamma\})$, through the pointwise definition:

$$\mathbf{IF}(\mu_w, f, S, \{\mu_y^\gamma\}) = \lim_{\gamma\to 0}\frac{S(f, \mu_y^\gamma) - S(f, \mu_y^0)}{\gamma}, \qquad f \in \left(-\frac{1}{2}, \frac{1}{2}\right].$$

The abbreviated notation $\mathbf{IF}_S(\mu_w, f)$ will be convenient. Denote the influence functional for the lag-l covariance $\mathrm{IF}_R(\mu_w, l) = \mathrm{IF}(\mu_w, R(l), \{\mu_y^\gamma\})$. Then

$$\mathbf{IF}_S(\mu_w, f) = \sum_{l=-\infty}^{\infty} \mathrm{IF}_R(\mu_w, l) e^{i2\pi f l}.$$

For the AO model with independent outliers of amplitude ξ, the influence curve for $R(l)$ is

$$\mathrm{IC}_R(\xi, l) = \begin{cases} \xi^2, & l = 0, \\ 0, & l \neq 0, \end{cases}$$

which gives the spectral density influence curve

$$\mathbf{IC}_S(\xi, f) = \xi^2, \qquad f \in \left(-\tfrac{1}{2}, \tfrac{1}{2}\right].$$

As one might expect, the influence curve in this case is quadratically unbounded, and independent of f. Other kinds of contamination—such as constant-level patches, sinusoidal patches, or patches with nonflat spectral structure—will lead to influence functionals and influence curves for $S(f)$ which depend on f in a nontrivial way, and are quadratically unbounded in the amplitude of contamination. It will be interesting to compute $\mathbf{IC}_S$ for these and other types of contaminations motivated by applications.

Robust alternatives to the smoothed periodogram estimates have been proposed in Kleiner, Martin, and Thomson (1979) and Martin and Thomson (1982). We intend to study these estimates in terms of their influence functionals and curves, which will probably have to be computed via simulation.

9. Proofs of theorems.

PROOF OF THEOREM 4.1. According to the definition of a $\tilde{\mathbf{\Psi}}$ estimate

$$\mathbf{m}\left(\gamma, \mathbf{T}\left(\mu_y^\gamma\right)\right) = 0.$$

By assumption (a) there exists $\gamma_0 > 0$ such that $|\mathbf{T}(\mu_y^\gamma) - \mathbf{t}_0| \leq \varepsilon$ for $\gamma \leq \gamma_0$, where ε is as in assumption (b). Therefore by (b) and the mean-value theorem we have for $\gamma \leq \min(\gamma_0, \varepsilon)$,

$$\mathbf{m}(\gamma, \mathbf{t}_0) + \mathbf{D}(\gamma, \mathbf{t}^*(\gamma))\left(\mathbf{T}\left(\mu_y^\gamma\right) - \mathbf{T}(\mu_x)\right) = 0$$

and, by assumption (a), $\mathbf{t}^*(\gamma) \to \mathbf{t}_0$ as $\gamma \to 0$. Then by assumption (b), $\mathbf{D}(\gamma, \mathbf{t}^*(\gamma)) \to \mathbf{C} = \mathbf{D}(0, \mathbf{t}_0)$ as $\gamma \to 0$. Then (4.6) follows. □

PROOF OF THEOREM 4.2. (i) Patches of length k are generated by letting

$$z_i^p = \max\left(\tilde{z}_i^p, \tilde{z}_{i-1}^p, \ldots, \tilde{z}_{i-k+1}^p\right),$$

where the $\tilde{z}_i^p$ are i.i.d. Bernoulli random variables with $P(\tilde{z}_1^p = 1) = p$, and

$\gamma = kp$, with k fixed. For fixed $m > 1$, put

$$C_{j,m}^{k,p} = \left\{\tilde{z}_{1-j}^{p} = 1\right\} \cap \left[\bigcap_{\substack{i=0\\ i\neq j}}^{m-1} \left\{\tilde{z}_{1-i}^{p} = 0\right\}\right], \qquad 0 \le j \le m-1, \tag{9.1}$$

$$C_{m,m}^{k,p} = \bigcap_{i=0}^{m-1} \left\{\tilde{z}_{1-i}^{p} = 0\right\}, \tag{9.2}$$

$$C_{m+1,m}^{k,p} = \bigcup_{\substack{i,j=0\\ i\neq j}}^{m-1} \left[\left\{\tilde{z}_{1-k}^{p} = 1\right\} \cap \left\{\tilde{z}_{1-j}^{p} = 1\right\}\right]. \tag{9.3}$$

Then for $0 \le j \le m-1$ we have

$$P\left(C_{j,m}^{k,p}\right) = p + d_{1,m}(p) = p + o(p). \tag{9.4}$$

We also have

$$P\left(C_{m,m}^{k,p}\right) = 1 + d_{2,m}(p) = 1 + O(p) \tag{9.5}$$

and

$$P\left(C_{m+1,m}^{k,p}\right) = d_{3,m}(p) = o(p). \tag{9.6}$$

According to the definition of $\mathbf{T}$ we have

$$E\tilde{\Psi}\left(\mathbf{y}_1^{\gamma}, \mathbf{T}\left(\mu_y^{\gamma}\right)\right) = 0.$$

Then

$$E\tilde{\Psi}(\mathbf{y}_1^{\gamma}, \mathbf{t}_0) + E\left[\Psi\left(\mathbf{y}_1^{\gamma}, \mathbf{T}\left(\mu_y^{\gamma}\right)\right) - \Psi(\mathbf{y}_1^{\gamma}, \mathbf{t}_0)\right] = 0.$$

Since $C_{j,m}^{k,p}$, $0 \le j \le m+1$, is a partition of the sample space of $(\tilde{z}_1^p, \tilde{z}_0^p, \dots, \tilde{z}_{2-m}^p)$ we have

$$E\tilde{\Psi}(\mathbf{y}_1^{\gamma}, \mathbf{t}_0) + \sum_{j=0}^{m+1} \mathbf{c}_j(m,p) P\left(C_{j,m}^{k,p}\right) = 0, \tag{9.7}$$

where

$$\mathbf{c}_j(m,p) = E\left[\tilde{\Psi}\left(\mathbf{y}_1^{\gamma}, \mathbf{T}\left(\mu_y^{\gamma}\right)\right) - \tilde{\Psi}(\mathbf{y}_1^{\gamma}, \mathbf{t}_0) | C_{j,m}^{k,p}\right]. \tag{9.8}$$

For $j \ge 0$, let $\mathbf{y}_{1,j} = (y_{1,j}, y_{0,j}, \dots)$ be given by

$$\mathbf{y}_{1,j} = \begin{cases} \left(\mathbf{w}_1^{1-j}, \mathbf{x}_{-j}\right), & 0 \le j \le k-1, \\ \left(\mathbf{x}_1^{1-j+k}, \mathbf{w}_{-j+k}^{1-j}, \mathbf{x}_{-j}\right), & j \ge k. \end{cases} \tag{9.9}$$

Conditioned on $C_{j,m}^{k,p}$, for $0 \le j \le m-1$ we have $y_{1-i}^{\gamma} = y_{1-i,j}$. For $0 \le i \le m-k$ and for $i > m-k$, y_{1-i}^{γ} is either x_{1-i} or w_{1-i}. Thus we have

$$\left|E\left[\tilde{\Psi}(\mathbf{y}_1^{\gamma}, \mathbf{t}) | C_{j,m}^{k,p}\right] - E\tilde{\Psi}(\mathbf{y}_{1,j}, \mathbf{t})\right| \le H_{m-k+1}(\mathbf{t}), \qquad 0 \le j \le m-1. \tag{9.10}$$

Then

$$|\mathbf{c}_j(m,p)| \le \left| E\left[\tilde{\Psi}\left(\mathbf{y}_{1,j},\mathbf{T}(\mu_y^\gamma)\right) - \tilde{\Psi}(\mathbf{y}_{1,j},\mathbf{t}_0)\right]\right| + H_{m-k+1}\left(\mathbf{T}(\mu_y^\gamma)\right) + H_{m-k+1}(\mathbf{t}_0).$$

Therefore by assumptions (a) and (e), and the dominated convergence theorem, we get

$$|\mathbf{c}_j(m,p)| \le b_j(m,p) + H_{m-k+1}\left(\mathbf{T}(\mu_y^\gamma)\right) + H_{m-k+1}(\mathbf{t}_0), \quad 0 \le j \le m-1, \tag{9.11}$$

where

$$\lim_{p\to 0} b_j(m,p) = 0. \tag{9.12}$$

We will now show that for $|\mathbf{t}-\mathbf{t}_0| \le \varepsilon_0$, where ε_0 is as in assumption (c), we have

$$\left| E\left[\tilde{\Psi}(\mathbf{y}_1^\gamma,\mathbf{t}) | C_{m,m}^{k,p}\right] - E\tilde{\Psi}(\mathbf{x}_1,\mathbf{t})\right| \le p \sum_{i=m-k+1}^{\infty} H_i(\mathbf{t}). \tag{9.13}$$

We have

$$E\left[\tilde{\Psi}(\mathbf{y}_1^\gamma,\mathbf{t}) | C_{m,m}^{k,p}\right] = (1-p)E\left[\tilde{\Psi}(\mathbf{y}_1^\gamma,\mathbf{t}) | C_{m,m}^{k,p}, z_{1-m}^p = 0\right] + pE\left[\tilde{\Psi}(\mathbf{y}_1^\gamma,\mathbf{t}) | C_{m,m}^{k,p}, z_{1-m}^p = 1\right].$$

Since conditionally on $C_{m,m}^{k,p}$, $y_{1-i}^\gamma = x_{1-i}$, $0 \le i \le m-k$, we have

$$\left| E\left[\tilde{\Psi}(\mathbf{y}_1^\gamma,\mathbf{t}) | C_{m,m}^{k,p}, z_{1-m}^p = 1\right] - E\tilde{\Psi}(\mathbf{x}_1,\mathbf{t})\right| \le H_{m-k+1}(\mathbf{t}).$$

Then

$$\begin{aligned}
&\left| E\left[\tilde{\Psi}(\mathbf{y}_1^\gamma,\mathbf{t}) | C_{m,m}^{k,p}\right] - E\tilde{\Psi}(\mathbf{x}_1,\mathbf{t})\right| \\
&\quad \le (1-p)\left| E\left[\tilde{\Psi}(\mathbf{y}_1^\gamma,\mathbf{t} | C_{m+1,m+1}^{k,p}\right] - E\tilde{\Psi}(\mathbf{x}_1,\mathbf{t})\right| \\
&\qquad + p\left| E\left[\tilde{\Psi}(\mathbf{y}_1^\gamma,\mathbf{t} | C_{m,m}^{k,p}, z_{1-m}^p = 1\right] - E\tilde{\Psi}(\mathbf{x}_1,\mathbf{t})\right| \\
&\quad \le \left| E\left[\tilde{\Psi}(\mathbf{y}_1^\gamma,\mathbf{t}) | C_{m+1,m+1}^{k,p}\right] - E\tilde{\Psi}(\mathbf{x}_1,\mathbf{t})\right| + pH_{m-k+1}(\mathbf{t}).
\end{aligned}$$

Iterating this relationship, we get

$$\begin{aligned}
&\left| E\left[\tilde{\Psi}(\mathbf{y}_1^\gamma,\mathbf{t}) | C_{m,m}^{k,p}\right] - E\tilde{\Psi}(\mathbf{x}_1,\mathbf{t})\right| \\
&\quad \le p \sum_{i=m-k+1}^{m-k+h+1} H_i(\mathbf{t}) + \left| E\left[\tilde{\Psi}(\mathbf{y}_1^\gamma,\mathbf{t}) | C_{m+h+1,m+h+1}^{k,p}\right] - E\tilde{\Psi}(\mathbf{x}_1,\mathbf{t})\right|.
\end{aligned}$$

Since the second term of the right-hand side is no greater than $H_{m+h-k+2}(\mathbf{t})$, which by (c) tends to 0 for $|\mathbf{t}-\mathbf{t}_0| \le \varepsilon_0$, we get (9.13). Using (9.13) we get

$$\begin{aligned}
&\left| \mathbf{c}_m(m,p) - E\left[\tilde{\Psi}\left(\mathbf{x}_1,\mathbf{T}(\mu_y^\gamma)\right) - \tilde{\Psi}(\mathbf{x}_1,\mathbf{t}_0)\right]\right| \\
&\quad \le p \sum_{i=m-k+1}^{\infty} \left[H_i\left(\mathbf{T}(\mu_y^\gamma)\right) + H_i(\mathbf{t}_0)\right].
\end{aligned}$$

Then according to (a) and (b) we get

$$\begin{aligned} &\left|\mathbf{c}_m(m,p) - \mathbf{C}\left[\mathbf{T}(\mu_y^\gamma) - \mathbf{T}(\mu_x)\right]\right| \\ &\quad \le a(p) + p \sum_{i=m-k+1}^{\infty} \left[H_i\left(\mathbf{T}(\mu_y^\gamma)\right) + H_i(\mathbf{t}_0)\right], \end{aligned} \tag{9.14}$$

where

$$a(p) = o(p). \tag{9.15}$$

We also have

$$|\mathbf{c}_{m+1}(m,p)| \le H_0\left(\mathbf{T}(\mu_y^\gamma)\right) + H_0(\mathbf{t}_0). \tag{9.16}$$

Using (2.5), (9.4)–(9.7), (9.11)–(9.12), and (9.14)–(9.16), along with assumptions (a), (c), and (d), straightforward computations give (see Martin and Yohai, 1984a, for details):

$$\left|\frac{E\tilde{\Psi}(\mathbf{y}_1^\gamma,\mathbf{t}_0)}{g(\gamma)} + \frac{\mathbf{C}\left[\mathbf{T}(\mu_y^\gamma) - \mathbf{T}(\mu_x)\right]}{g(\gamma)}\right| = O(1)$$

and thereby (4.6) follows.

(ii) Using (9.1), (9.2), and (9.3), we have

$$E\tilde{\Psi}(\mathbf{y}_1^\gamma,\mathbf{t}_0) = \sum_{j=0}^{m+1} E\left[\tilde{\Psi}(\mathbf{y}_1^\gamma,\mathbf{t}_0)|C_{j,m}^{k,p}\right] P\left(C_{j,m}^{k,p}\right).$$

Then using (9.4), (9.5), and (9.6) we get

$$\begin{aligned} &\left|\frac{E\tilde{\Psi}(\mathbf{y}_1^\gamma,\mathbf{t}_0)}{g(\gamma)} - \frac{1}{k}\sum_{j=0}^{\infty}\mathbf{G}_j^k\right| \\ &\quad \le \frac{p + d_{1,m}(p)}{g(\gamma)} \sum_{j=0}^{m-1} \left|E\left[\tilde{\Psi}(\mathbf{y}_1^\gamma,\mathbf{t}_0)|C_{j,m}^{k,p}\right] - \mathbf{G}_j^k\right| \\ &\quad\quad + \left|\frac{1}{k} - \frac{p + d_{1,m}(p)}{g(\gamma)}\right| \sum_{j=0}^{m-1} |\mathbf{G}_j^k| + \frac{1}{k}\sum_{j=m}^{\infty}|\mathbf{G}_j^k| \\ &\quad\quad + \frac{1 + d_{2,m}(p)}{g(\gamma)} E\left[\tilde{\Psi}(\mathbf{y}_1^\gamma,\mathbf{t}_0)|C_{m,m}^{k,p}\right] \\ &\quad\quad + \frac{d_{3,m}(p)}{g(\gamma)} E\left[\tilde{\Psi}(\mathbf{y}_1^\gamma,\mathbf{t}_0)|C_{m+1,m}^{k,p}\right]. \end{aligned} \tag{9.17}$$

Since $\mathbf{G}_j^k = E\tilde{\Psi}(\mathbf{y}_{1,j},\mathbf{t}_0)$, $0 \le j \le m-1$, where $\mathbf{y}_{1,j}$ is given by (9.9), using (9.10) gives

$$\left|E\left[\tilde{\Psi}(\mathbf{y}_1^\gamma,\mathbf{t}_0)|C_{j,m}^{k,p}\right] - \mathbf{G}_j^k\right| \le H_{m-k+1}(\mathbf{t}_0), \qquad 0 \le j \le m-1. \tag{9.18}$$

We also have

$$E\left[\tilde{\Psi}(\mathbf{y}_1^\gamma, \mathbf{t}_0) | C_{m+1, m}^{k, p}\right] \le H_0(\mathbf{t}_0). \tag{9.19}$$

Since $E\tilde{\Psi}(\mathbf{x}_1, \mathbf{t}_0) = 0$, and since $\mathbf{x}_1$ and $\mathbf{y}_{1,j}$ have the same first $j - k + 1$ components when $j \ge k$, we have

$$|\mathbf{G}_j^k| \le H_{j-k+1}(\mathbf{t}_0), \qquad j \ge k, \tag{9.20}$$

and by the definitions of $\mathbf{G}_j^k$ and $H_0(\mathbf{t})$ we have

$$|\mathbf{G}_j^k| \le H_0(t_0), \qquad 0 \le j \le k - 1. \tag{9.20$'$}$$

Conditioned on $C_{m,m}^{k,p}$, $\mathbf{x}_1$ and $\mathbf{y}_1^\gamma$ have the same first $m - k + 1$ components, and so (9.13) gives

$$\left|E\left[\tilde{\Psi}(\mathbf{y}_1^\gamma, \mathbf{t}_0) | C_{m,m}^{k,p}\right]\right| \le p \sum_{i=m-k+1}^{\infty} H_i(\mathbf{t}_0). \tag{9.21}$$

Using (2.5), (9.4)–(9.6), and (9.18)–(9.21), along with assumptions (a), (c), and (d), we get (4.8) by straightforward computation (again, details may be found in Martin and Yohai, 1984a). □

PROOF OF THEOREM 4.3. By (4.6), (b), (c), and the dominated convergence theorem, we have

$$\mathbf{IF}\left(\mu_w, \mathbf{T}, \{\mu_y^\gamma\}\right) = \lim_{\gamma \to 0} \mathbf{C}^{-1} \sum_{j=1}^{\infty} \frac{E\pi_j\left(\mathbf{y}_{1,I_j}^\gamma, \mathbf{t}_0\right)}{g(\gamma)}. \tag{9.22}$$

Note that

$$\begin{aligned} E\pi_j\left(\mathbf{y}_{1,I_j}^\gamma, \mathbf{t}_0\right) &= E\left[\pi_j\left(\mathbf{x}_{1,I_j}, \mathbf{t}_0\right) | \mathbf{z}_{1,I_j}^\gamma = \mathbf{0}\right]\left(1 - g_j^*(\gamma)\right) \\ &\quad + E\left[\pi_j\left(\mathbf{y}_{1,I_j}^\gamma, \mathbf{t}_0\right) | \mathbf{z}_{1,I_j}^\gamma \ne \mathbf{0}\right] g_j^*(\gamma), \end{aligned} \tag{9.22$'$}$$

where $g_j^*(\gamma) = P(\mathbf{z}_{1,I_j}^\gamma \ne \mathbf{0})$. Also

$$\begin{aligned} 0 = E\pi_j\left(\mathbf{x}_{1,I_j}, \mathbf{t}_0\right) &= E\left[\pi_j\left(\mathbf{x}_{1,I_j}, \mathbf{t}_0\right) | \mathbf{z}_{1,I_j}^\gamma = \mathbf{0}\right]\left(\left(1 - g_j^*(\gamma)\right)\right) \\ &\quad + E\left[\pi_j\left(\mathbf{x}_{1,I_j}, \mathbf{t}_0\right) | \mathbf{z}_{1,I_j}^\gamma \ne \mathbf{0}\right] g_j^*(\gamma) \end{aligned}$$

Then we have

$$\left|E\left[\pi_j\left(\mathbf{x}_{1,I_j}, \mathbf{t}_0\right) | \mathbf{z}_{1,I_j}^\gamma = \mathbf{0}\right]\right| \le \frac{K_j g_j^*(\gamma)}{1 - g_j^*(\gamma)}.$$

Since $g_j^*(\gamma) \le hg(\gamma)$, using (9.22′) we have

$$\left|E\pi_j\left(\mathbf{y}_{1,I_j}^\gamma, \mathbf{t}_0\right)\right| \le K_j hg(\gamma)\left[\frac{1}{1 - hg(\gamma)} + 1\right]$$

for sufficiently small γ, and so (2.1) and (9.22) give (4.9). □

PROOF OF THEOREM 5.1. (i) For the GM estimates, straightforward computations show that

$$C = E\left\{\frac{\partial}{\partial \phi}\tilde{\psi}(\mathbf{x}_1, \phi)\right\} = -\frac{B}{(1-\phi^2)^{1/2}}. \tag{9.23}$$

For the RA estimates

$$C = -\frac{B}{1-\phi^2} \tag{9.24}$$

and the result (i) follows from (6.23) and (6.24).

(ii) The result is proved in Bustos (1982) for the GM estimates, and in Bustos, Fraiman, and Yohai (1984) for RA estimates. □

PROOF OF THEOREM 5.2. Using (A1)–(A4) and $E|v_1|^{k_1+k_2} < \infty$, it is easy to show that there exist solutions $t_\gamma^{\mathrm{GM}} = T^{\mathrm{GM}}(\mu_y^\gamma)$ and $t_\gamma^{\mathrm{RA}} = T^{\mathrm{RA}}(\mu_y^\gamma)$ of

$$E\eta\left(y_1^\gamma - t y_0^\gamma, (1-t^2)^{1/2} y_0^\gamma\right) = 0$$

and

$$\sum_{j=1}^{\infty} t^{j-1} E\eta\left(y_1^\gamma - t y_0^\gamma, y_{1-j}^\gamma - t y_{-j}^\gamma\right) = 0,$$

respectively, such that $t_\gamma^{\mathrm{GM}} \to \phi$ and $t_\gamma^{\mathrm{RA}} \to \phi$ as $\gamma \to 0$. Then assumption (a) of Theorem 4.2 is satisfied. It is easy to check that the other assumptions of Theorem 4.2 are also satisfied, and so T^{GM} and T^{RA} satisfy (4.8).

Let $\mathbf{y}_{1,j} = (y_{1,j}, y_{0,j}, \dots)$ be given by

$$\mathbf{y}_{1,j} = \begin{cases} \left(\mathbf{x}_1^{1-j} + \mathbf{v}_1^{1-j}, \mathbf{x}_{-j}\right), & \text{if } 0 \le j \le k-1, \\ \left(\mathbf{x}_1^{1-j+k}, \mathbf{x}_{-j+k}^{1-j} + \mathbf{v}_{-j+k}^{1-j}, \mathbf{x}_{-j}\right), & \text{if } j \ge k. \end{cases}$$

Then

$$G_j^k = E\tilde{\psi}(\mathbf{y}_{1,j}, \phi).$$

For GM estimates $\tilde{\psi}(\mathbf{y}_1, \phi) = \eta(y_1 - \phi y_0, y_0(1-\phi^2)^{1/2})$, and so we get

$$G_j^k = \begin{cases} 0, & j = 0, \\ E\eta\left(u_1 + v_1 - \phi v_0, (x_0 + v_0)(1-\phi^2)^{1/2}\right), & j \le k-1, \\ E\eta\left(u_1 - \phi v_0, (x_0 + v_0)(1-\phi^2)^{1/2}\right), & j = k, \\ 0, & j > k. \end{cases}$$

Therefore by (4.8) and (9.23) we have

$$\mathrm{IF}_{\mathrm{AO},R}\left(\mu_v, T^{\mathrm{GM}}, \phi\right) = \frac{(1-\phi^2)^{1/2}}{kB}\left[(k-1)G_1^k + G_k^k\right],$$

which gives (5.8), with (5.6) as a special case.

Now for T^{RA}, $\tilde{\psi}(\mathbf{y}_1, \phi)$ is given by (3.13), and so

$$G_j^k = \sum_{h=1}^{\infty} \phi^{h-1} G_{j,h}^k,$$

with

$$G_{j,h}^k = E\eta\left(y_{1,j} - \phi y_{0,j}, y_{1-h,j} - \phi y_{-h,j}\right).$$

Using the assumptions on η gives

$$G_{j,h}^k = \begin{cases} 0, & h > j, \\ 0, & j > k, \\ E\eta\left(u_1 - \phi v_0, u_{1-k} + v_{1-k}\right), & h = j = k, \\ E\eta\left(u_1 - \phi v_0, u_{1-h} + v_{1-h} - \phi v_{-h}\right), & h > j = k, \\ E\eta\left(u_1 + v_1 - \phi v_0, u_{1-h} + v_{1-h}\right), & h = j < k, \\ E\eta\left(u_1 + v_1 - \phi v_0, u_{1-h} + v_{1-h} - \phi v_{-h}\right), & h < j < k. \end{cases}$$

Therefore by (4.8) and (9.24) we have

$$\mathrm{IF}\left(\mu_v, T^{\mathrm{GM}}, \phi\right) = \frac{1-\phi^2}{kB}\left[\sum_{h=1}^{k-2} (k-h-1)\phi^{h-1} G_{k-1,h}^k + \sum_{h=1}^{k-1} \phi^{h-1} G_{k-1,k-1}^k + \sum_{h=1}^{k-1} \phi^{h-1} G_{k,h}^k + \phi^{k-1} G_{k,k}^k\right].$$

This gives (5.9), with (5.7) as a special case. □

PROOF OF THEOREM 5.3. (i) Straightforward computations show that for T^{GM} we have $C = B/(1-\theta^2)^{1/2}$, and for T^{RA} we have $C = B/(1-\theta^2)$. Then (5.10) and (5.11) follow.

(ii) The result is proved in Bustos, Fraiman, and Yohai (1984) for RA estimates. For GM estimates, it can be proved similarly. □

PROOF OF THEOREM 5.4. (A1)–(A4) and $E|\nu_1|^{k_1+k_2} < \infty$ imply as in Theorem 5.2 the existence of solutions $T^{\mathrm{GM}}(\mu_y^\gamma)$ and $T^{\mathrm{RA}}(\mu_y^\gamma)$ which converge to the true parameter θ as $\gamma \to 0$. Then assumption (a) of Theorem 4.2 holds. We will show that assumption (c) is satisfied. We will assume $h_1 = h_2 = 0$, but the proof in the other cases is similar.

For GM estimates we have

$$\tilde{\psi}(\mathbf{y}_1, \theta) = \eta\left(\sum_{i=0}^{\infty} \theta^i y_{1-i}, (1-\theta^2)^{1/2} \sum_{i=0}^{\infty} (1+i)\theta^i y_{-i}\right).$$

Straightforward computations show that

$$H_m(\theta) \le 2KM\left[\frac{|\theta|^m}{1-|\theta|} + \left(\frac{m|\theta|^{m-1}}{1-|\theta|} + \frac{|\theta|^m}{(1-|\theta|)^2}\right)(1-\theta^2)^{1/2}\right],$$

where K is a bound on the partial derivatives of η, and $M = E|x_1| + E|v_1|$. Therefore assumption (c) is satisfied for T^{GM}. For the RA estimates

$$\tilde{\psi}(\mathbf{y}_1, \theta) = \sum_{j=1}^{\infty} \theta^{j-1}\eta\left(\sum_{i=0}^{\infty} \theta^i y_{1-i}, \sum_{i=0}^{\infty} \theta^i y_{1-j-i} \right),$$

and straightforward computations show that

$$H_m(\theta) \le \frac{2KM|\theta|^m}{(1-|\theta|)^2} + \frac{2KMm|\theta|^{m-1}}{1-|\theta|} + \frac{2KM|\theta|^m}{(1-|\theta|)^2}.$$

Thus assumption (c) of Theorem 4.2 is satisfied for T^{RA}. It is easy to check the remaining assumptions of Theorem 4.2 for both T^{GM} and T^{RA}. Therefore (4.8) holds for these estimates.

Now we need to evaluate (4.8) for T^{GM} and T^{RA}. For GM estimates

$$G_j^1 = E\eta\left(u_1 + \theta^j v_{1-j}, (1-\theta^2)^{1/2}\left(s_0 + j\theta^{j-1} v_{1-j}\right)\right).$$

Since $(1-\theta^2)^{1/2} s_0$ is $N(0,1)$ and independent of u_1, and $C = B/(1-\theta^2)^{1/2}$, (4.8) gives (5.13). For RA estimates we have

$$G_j^1 = \sum_{h=1}^{j} \theta^{h-1} E\eta\left(u_1 + \theta^j v_{1-j}, u_{1-h} + \theta^{j-h} v_{1-h}\right)$$
$$+ \sum_{h=j+1}^{\infty} \theta^{h-1} E\eta\left(u_1 + \theta^j v_{1-j}, u_{1-h}\right).$$

By (A1), $E\eta(u_1 + \theta^j v_{1-j}, u_{1-h}) = 0$. Since $C = B/(1-\theta^2)$, interchanging the order of summations gives (5.14). □

Proof of Theorem 7.1. (i) It is easy to check that when $y_i \equiv x_i$ is an AR(1) process, the V_l in (4.4) are zero for $l > 0$. For $\tilde{\psi}$ satisfying Theorem 4.2 we have

$$\mathrm{ICH}(\mathbf{y}_1) = \mathrm{ICH}(\mathbf{x}_1, \phi) = C^{-1}\tilde{\psi}(\mathbf{x}_1, \phi)$$

with $C = C_{\tilde{\psi}}$ given in (4.2′), and so from (4.4) we have

$$V_0(\tilde{\psi}) = C_{\tilde{\psi}}^{-2} E\tilde{\psi}^2(\mathbf{x}_1, \phi).$$

Our proof consists of showing that $C_{\tilde{\psi}} = C_{\tilde{\psi}^*}$, and that $E\tilde{\psi}^{*2}(\mathbf{x}_1, \phi) \le E\tilde{\psi}^2(\mathbf{x}_1, \phi)$, which implies $V_0(\tilde{\psi}^*) \le V_0(\tilde{\psi})$. We first show that

$$E\tilde{\psi}^{*2}(\mathbf{x}_1, \phi) \le E\tilde{\psi}^2(\mathbf{x}_1, \phi). \tag{9.25}$$

In order to prove (9.25) it is enough to show that

$$E\left[\tilde{\psi}(\mathbf{x}_1, \phi) - \tilde{\psi}^*(\mathbf{x}_1, \phi)\right]\tilde{\psi}^*(\mathbf{x}_1, \phi) = 0.$$

Let

$$\mathbf{u}_1 = \mathbf{u}_1(\phi) = (u_1, u_0, u_{-1}, \dots) = (u_1(\phi), u_0(\phi), u_{-1}(\phi), \dots).$$

Then in view of the definition of $\tilde{\psi}^*$ and $\tilde{\psi}^+$, it suffices to show that

$$E\tilde{\psi}^*(\mathbf{x}_1, \phi)\eta_i(u_1, u_{1-i}, \phi) = E\tilde{\psi}^+(\mathbf{u}_1, \phi)\eta_i(u_1, \mathbf{u}_{1-i}, \phi), \quad i = 1, 2, \dots . \tag{9.26}$$

But

$$\begin{aligned} E\tilde{\psi}^+(\mathbf{u}_1, \phi)\eta_i(u_1, u_{1-i}, \phi) &= E\eta_i(u_1, u_{1-i}, \phi)E\left[\tilde{\psi}^+(\mathbf{u}_1, \phi)|u_1, u_{1-i}\right] \\ &= E\eta_i^2(u_1, u_{1-i}, \phi), \end{aligned}$$

and since (C3) implies

$$E\eta_i(u_1, u_{1-i}, \phi)\eta_j(u_1, u_{1-j}, \phi) = 0, \qquad i \neq j,$$

we have

$$E\tilde{\psi}^*(\mathbf{x}_1, \phi)\eta_i(u_1, u_{1-i}, \phi) = E\eta_i^2(u_1, u_{1-i}, \phi),$$

from which (9.26) follows.

Now we will establish that $C_{\tilde{\psi}} = C_{\tilde{\psi}^*}$, that is,

$$E\mathrm{D}\tilde{\psi}(\mathbf{x}_1, \phi) = E\mathrm{D}\tilde{\psi}^*(\mathbf{x}_1, \phi),$$

where $\mathrm{D} = \partial/\partial\phi$. We have

$$\mathrm{D}\tilde{\psi}(\mathbf{x}_1, \phi) = -\sum_{i=1}^{\infty} x_{1-i}\tilde{\psi}_i^+(\mathbf{u}_1(\phi), \phi) + \tilde{\psi}_\phi^+(\mathbf{u}_1(\phi), \phi),$$

where

$$\tilde{\psi}_i^+(a_1, a_2, \dots, \phi) = (\partial/\partial a_i)\tilde{\psi}^+(a_1, a_2, \dots, \phi)$$

and

$$\tilde{\psi}_\phi^+(a_1, a_2, \dots, \phi) = (\partial/\partial\phi)\tilde{\psi}^+(a_1, a_2, \dots, \phi).$$

By (C2) we have

$$Ex_{1-i}\tilde{\psi}_i^+(\mathbf{u}_1, \phi) = 0, \qquad i > 1,$$

and

$$E\psi_\phi^+(\mathbf{u}_1, \phi) = 0.$$

Therefore,

$$E\mathrm{D}\tilde{\psi}(\mathbf{x}_1, \phi) = -Ex_0\tilde{\psi}_1^+(\mathbf{u}_1, \phi) = -\sum_{i=1}^{\infty}\phi^{i-1}Eu_{1-i}\tilde{\psi}_1^+(\mathbf{u}_1, \phi).$$

Since (C1) gives

$$\begin{aligned} Eu_{1-i}\tilde{\psi}_1^+(u_1, u_0, \dots, \phi) &= Eu_{1-i}E\left[\tilde{\psi}_1^+(u_1, u_0, \dots, \phi)|u_1, u_{1-i}\right] \\ &= Eu_{1-i}(\partial/\partial u_1)\eta_i(u_1, u_{1-i}, \phi), \end{aligned}$$

we have

$$E\mathrm{D}\tilde{\psi}(\mathbf{x}_1, \phi) = -\sum_{i=1}^{\infty}\phi^{i-1}Eu_{i-1}(\partial/\partial u_1)\eta_i(u_1, u_{1-i}, \phi).$$

Now similar reasoning gives exactly the same expression for $E\mathrm{D}\tilde{\psi}^*(\mathbf{x}_1, \phi)$, and the proof of (i) is complete.

(ii) By Theorem 4.2 we have

$$\mathrm{IF}_{\mathrm{AO},1}(\mu_v, T, \phi) = -\frac{1}{C}\sum_{i=0}^{\infty} G_i^1,$$

where

$$C = E\mathrm{D}\tilde{\psi}(\mathbf{x}_1, \phi),$$

$$G_0^1 = E\tilde{\psi}^+(u_1 + v_1, u_0, u_{-1}, \dots, \phi),$$

and for $i \geq 1$,

$$G_i^1 = E\tilde{\psi}^+(u_1, \dots, u_{3-i}, u_{2-i} - \phi v_{1-i}, u_{1-i} + v_{1-i}, u_{-i}, \dots, \phi).$$

(C2) implies $G_0^1 = 0$ and (C3) implies $G_i^1 = 0$ for $i \geq 2$. Therefore

$$\mathrm{IF}_{\mathrm{AO},1}(\mu_v, T, \phi) = -\frac{1}{C}G_1^1 = -\frac{1}{C}E\eta_1(u_1 - \phi v_0, u_0 + v_0, \phi).$$

It is easy to check that for T^*

$$G_i^1 = \begin{cases} 0, & \\ E\eta_1(u_1 - \phi v_0, u_0 + v_0, \phi), & i = 0 \text{ and } i \geq 2 \end{cases}$$

and so one also has

$$\mathrm{IF}_{\mathrm{AO},1}(\mu_v, T^*, \phi) = -\frac{1}{C}E\eta_1(u_1 - \phi v_0, u_0 + v_0, \phi). \qquad \square$$

Acknowledgments. The authors appreciate the help of Judith Zeh, who wrote the code for computing the influence curves displayed in Figures 1–3. She also prepared the final plots, using the interactive statistical language and system S (Becker and Chambers, 1984). We wish to thank the Institute of Pure and Applied Mathematics (IMPA) in Rio de Janeiro for their generous support for two months of 1983, when this research was in its incipient stages. We also would like to thank the referees of the paper for very thorough and illuminating reviews.

REFERENCES

ANDREWS, D. F., BICKEL, P. J., HAMPEL, F. R., HUBER, P. J., ROGERS, W. H. and TUKEY, J. W. (1972). *Robust Estimates of Location-Survey and Advances.* Princeton Univ. Press, Princeton, N.J.

BECKER, R. A. and CHAMBERS, J. M. (1984). *S: An Interactive Environment for Data Analysis and Graphics.* Wadsworth, Belmont, Calif.

BUSTOS, O. (1982). General M-estimates for contaminated pth-order autoregressive processes: consistency and asymptotic normality. *Z. Wahrsch. verw. Gebiete* **59** 491–504.

BUSTOS, O., FRAIMAN, R. and YOHAI, V. J. (1984). Asymptotic behavior of the estimates based on residual autocovariances for ARMA models. In *Robust and Nonlinear Time Series Analysis* (J. Franke, W. Härdle, and D. Martin, eds.) 26–49. Springer, New York.

BUSTOS, O. and YOHAI, V. J. (1986). Robust estimates for ARMA models. *J. Amer. Statist. Assoc.* **81** 155–168.

CHERNICK, M. R., DOWNING, D. J. and PIKE, D. H. (1982). Detecting outliers in time series data. *J. Amer. Statist.* **77** 743–747.

DENBY, L. and MARTIN, R. D. (1979). Robust estimation of the first order autoregressive parameter. *J. Amer. Statist. Assoc.* **74** 140–146.

HAMPEL, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69** 383–393.

HAMPEL, F. R. (1975). Beyond location parameters: robust concepts and methods. *Bull. Internat. Statist. Inst.* **40**(1) 375–382.

HAMPEL, F. R. (1978). Optimally bounding the gross-error-sensitivity and the influence of position in factor space. *Proc. ASA Statist. Computing Sec.* 59–64. Amer. Statist. Assoc., Washington.

HUBER, P. (1981). *Robust Statistics.* Wiley, New York.

HUBER, P. J. (1983). Minimax aspects of bounded influence regression. *J. Amer. Statist. Assoc.* **78** 66–72.

JAMES, B. R. and JAMES, K. L. (1983). On the influence curve for quantal bioassay. *J. Statist. Plann. Inference* **8** 331–345.

KELLY, G. E. (1984). The influence function in the errors in variables problem. *Ann. Statist.* **12** 87–100.

KLEINER, B., MARTIN, R. D. and THOMSON, D. J. (1979). Robust estimation of power spectra. *J. Roy. Statist. Soc. Ser. B* **41** 313–351.

KRASKER, W. S. and WELSCH, R. E. (1982). Efficient bounded-influence regression estimation. *J. Amer. Stat. Assoc.* **77** 595–604.

KÜNSCH, H. (1984). Infinitesimal robustness for autoregressive processes. *Ann. Statist.* **12** 843–863.

LAMBERT, D. (1981). Influence functions for testing. *J. Amer. Statist. Assoc.* **76** 649–657.

MALLOWS, C. L. (1976). On some topics in robustness. Bell Labs Technical Memo, Murray Hill, N.J. (Talks given at NBER Workshop on Robust Regression, Cambridge, Mass., May 1973, and at ASA-IMS Regional Meeting, Rochester, N.Y., May, 1975.)

MARTIN, R. D. (1980). Robust estimation of autoregressive models (with discussion). In *Directions in Time Series* (D. R. Brillinger and G. C. Tiao, eds.) 228–262. IMS, Hayward, Calif.

MARTIN, R. D. (1981). Robust methods for time series. In *Applied Time Analysis Series II* (D. F. Findley, ed.) 683–759. Academic, New York.

MARTIN, R. D. (1982). The Cramér–Rao bound and robust M-estimates for autoregressions. *Biometrika* **69** 437–442.

MARTIN, R. D. and JONG, J. (1977). Asymptotic properties of robust generalized M-estimates for the first-order autoregressive parameter. Bell Labs Technical Memo, Murray Hill, N.J.

MARTIN, R. D. and THOMSON, D. J. (1982). Robust-resistant spectrum estimation. *Proc. IEEE* **70** 1097–1115.

MARTIN, R. D. and YOHAI, V. J. (1984a). Influence curves for time series. Technical Report No. 51, Dept. Statistics, Univ. of Washington, Seattle.

MARTIN, R. D. and YOHAI, V. J. (1984b). Gross-error sensitivities of GM and RA-estimates. In *Robust and Nonlinear Time Series Analysis* (J. Franke, W. Härdle and D. Martin, eds.) 198–217. Springer, New York.

MARTIN, R. D. and YOHAI, V. J. (1985). Robustness in time series and estimating ARMA models. In *Handbook of Statistics* **5** (E. J. Hannan, P. R. Krishnaiah and M. M. Rao, eds.) 119–155. Elsevier, New York.

MICHAEL, J. R. and SCHUCANY, W. R. (1985). The influence curve and goodness of fit. *J. Amer. Statist. Assoc.* **80** 678–682.

PORTNOY, S. L. (1977). Robust estimation in dependent situations. *Ann. Statist.* **5** 22–43.

RONCHETTI, E. (1982). Robust testing in linear models: the infinitesimal approach. Ph.D. dissertation, ETH, Zurich.

SAMUELS, S. J. (1978). Robustness of survival estimators. Ph.D. dissertation, Dept. Biostatistics, Univ. of Washington, Seattle.

DEPARTMENT OF STATISTICS, GN-22
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF BUENOS AIRES
CUIDAD UNIVERSITARIA, PABELLON 1
1428 BUENOS AIRES
ARGENTINA