The Annals of Statistics
1987, Vol. 15, No. 3, 1293–1312

QUALITATIVE ROBUSTNESS FOR STOCHASTIC PROCESSES

BY GRACIELA BOENTE, RICARDO FRAIMAN AND VICTOR J. YOHAI[1]

Universidad de Buenos Aires and CONICET, Universidad de Buenos Aires, and Universidad de Buenos Aires and CEMA.

In this paper we generalize Hampel's concept of qualitative robustness of a sequence of estimators to the case of stochastic processes with non-i.i.d. observations, defining appropriate metrics between samples. We also present a different approach to qualitative robustness which formalizes the notion of resistance. We give two definitions based on this approach: strong and weak resistance. We show that for estimating a finite dimensional real parameter, π-robustness is equivalent to weak resistance and, in the i.i.d. case, is also equivalent to strong resistance. Finally, we prove the strong resistance of a class of estimators which includes common GM-estimates for linear models and autoregressive processes.

1. Introduction. Hampel (1971) introduced a definition of qualitative robustness of a sequence of estimators for the case of independent and identically distributed (i.i.d.) observations. This definition states that a sequence of estimators T_n is robust at a given distribution μ on the sample space X, if for any distribution ν close to μ in the Prohorov metric, the laws of T_n under μ and ν are close in the Prohorov metric, uniformly in the sample size.

The use of the Prohorov distance reflects the intuitive meaning of robustness as insensitivity of the estimator to

(a) small errors in all the observations (e.g., round-off errors),
(b) a small fraction of large observations (outliers).

Hampel also defines the more restrictive concept of π-robustness which also requires insensitivity to "small" non-i.i.d. *deviations* from a nominal i.i.d. model.

The generalization of these definitions to the case of stochastic processes with dependent observations requires defining appropriate distances between distributions on X^n in the case of π-robustness and between distributions on X^∞ in the case of robustness. There is no unique natural way of doing this, and so several definitions of qualitative robustness based on different metrics between probability measures on X^∞ have been given. See, for example, Papantoni-Kazakos and Gray (1979) and Cox (1981).

Cox's (1981) proposal is not completely general, since it only makes sense for estimators which depend only on finite dimensional marginal empirical distributions. Many estimates, such as the usual least-squares estimates for the parameters of moving-average processes, do not satisfy this requirement.

Received March 1983; revised January 1987.

[1]Research supported in part by National Science Foundation Grant SES80-15570 and by the Office of Naval Research under Contract N00014-82-K-0062 while this author was visiting the Department of Statistics of the University of Washington, Seattle.

AMS 1980 *subject classifications*. Primary 62F35; secondary 62M10.

Key words and phrases. Qualitative robustness, robust estimation, GM-estimators, stochastic processes, autoregressive models.

A shortcoming of the "rho-bar" metric, proposed by Papantoni-Kazakos and Gray (1979), which was mentioned by Cox (1981), is that this metric is not invariant with respect to equivalent metrics d on the sample X. Moreover, the definition of robustness based on the "rho-bar" metric only has a natural intuitive meaning when d is bounded. In fact, Cox (1981) shows that when d is the usual metric on R, the sample mean is robust with respect to the Papantoni-Kazakos and Gray definition.

We will give two simple examples of time series parametric models which show the relevance of a suitable definition of qualitative robustness and the special problems arising when the observations have a time dependence structure.

Stationary autoregressive process of order 1 $(AR(1))$. Let $x_1, \dots, x_T$ be observations satisfying $x_t = \phi x_{t-1} + u_t$, $1 \leq t \leq T$, where $|\phi| < 1$ and the u_t's are i.i.d. random variables with symmetric distribution F. If F is normal the least squares (LS) estimate is asymptotically optimal, i.e., its asymptotic variance attains the Cramér–Rao bound. The LS-estimate, denoted here by $\hat{\phi}_{\mathrm{LS}}$, is defined by minimizing $\Sigma_{t=2}^{T} u_T^2(\phi)$, where $u_t(\phi) = x_t - \phi x_{t-1}$ and is given by $\hat{\phi}_{\mathrm{LS}} = \Sigma_{t=2}^{T} x_t x_{t-1} / \Sigma_{t=2}^{T} x_{t-1}^2$.

It is easy to show that just one outlier may produce a large change in $\hat{\phi}_{\mathrm{LS}}$, e.g., if $x_1 \to \infty$ while all the other x_t are fixed, then $\hat{\phi}_{\mathrm{LS}} \to 0$. [Similar results hold for the general AR(p) case.] Therefore, the LS-estimate does not satisfy criterion (b) of the intuitive notion of robustness given earlier.

The class of M-estimates for autoregression is defined similarly as for regression [Huber (1973)]. Given an even and nondecreasing function ρ, the AR(1) M-estimate is defined by the value $\hat{\phi}_{\mathrm{M}}$ which minimizes $\Sigma_{t=2}^{T} \rho(u_t(\phi))$. Let $\psi = \rho'$, then $\hat{\phi}_{\mathrm{M}}$ satisfies $\Sigma_{t=2}^{T} \psi(u_t(\phi)) x_{t-1} = 0$. For the special case of location, M-estimates based on bounded ψ-functions are qualitatively robust [see Huber (1981), Chapter 3, page 52]. For perfectly observed autoregression M-estimates may be designed to have high efficiency robustness, even when the distribution F of the u_t has "heavy" tails which give rise to innovation outliers [see Denby and Martin (1979) and Martin (1982)]. However, it is easy to check that if ρ is convex and, therefore, $\psi(x)$ is nondecreasing, we will have again that, with $x_2, \dots, x_T$ fixed, $\lim_{x_1 \to \infty} \hat{\phi}_{\mathrm{M}} = 0$. Autoregression M-estimates are not at all robust toward general changes in the true model other than the rather special change in the innovations distribution F only.

In order to obtain robustness, Denby and Martin (1979), Martin (1982) and Bustos (1982) proposed using the class of generalized M-estimates (GM-estimates) for autoregressive models. See also Künsch (1984). These estimates, often called bounded influence estimates, were first proposed by Hampel (1975) and Mallows (1976) for regression. In the case of the AR(1) model, a GM-estimate is defined as a solution of the equation $\Sigma_{t=2}^{T} \eta(u_t(\phi), x_{t-1}) = 0$, where η: $R^2 \to R$ is a bounded and continuous function which is odd in each argument. Two interesting subclasses of η functions are the Hampel type, where $\eta(u, v) = \psi(u \cdot v)$, and the Mallows type, where $\eta(u, v) = \psi_1(u)\psi_2(v)$. We will show in Section 5 that under general regularity conditions GM-estimates satisfy a suitable definition of qualitative robustness.

Moving average process of order 1 (*MA*(1)). Set $x_t = -\theta u_{t-1} + u_t$, where $|\theta| < 1$ and the u_t's are i.i.d. random variables with symmetric distribution F. The M-estimates are defined by minimizing $\Sigma_{t=2}^T \rho(u_t(\theta))$ with $u_t(\theta) = x_t + \theta x_{t-1} + \cdots + \theta^{t-1} x_1$.

If $\rho(u) = u^2$ we have the LS-estimate which is asymptotically efficient when F is normal. Differentiating the M-estimate loss function gives the following equation for the M-estimate $\hat{\theta}_{\mathrm{M}}$: $\Sigma_{t=2}^T \psi(u_t(\theta)) a_{t-1}(\theta) = 0$, where $a_t(\theta) = x_t + 2\theta x_{t-1} + \cdots + t\theta^{t-1} x_1$.

As in the AR(1) model, it is easy to show that if ρ is convex, we have $\lim_{x_1 \to \infty} \hat{\theta}_{\mathrm{M}} = 0$, all other x_t being fixed, and, therefore, $\hat{\theta}_{\mathrm{M}}$ does not satisfy criterion (b) of the heuristic notion of robustness.

By analogy with the GM-estimates for the AR(1) model, we define a GM-estimate for the MA(1) model as a solution of the equation $\Sigma_{t=2}^T \eta(u_t(\theta), a_{t-1}(\theta)) = 0$. It is possible to show that even if η is bounded and also odd and monotone in each variable then $\lim_{x_1 \to \infty} \hat{\theta}_{\mathrm{GM}} = 0$ and, therefore, $\hat{\theta}_{\mathrm{GM}}$ may not be considered robust either. This fact, which is in striking contrast to the AR(1) case, may be explained by the fact that although each term $\eta(u_t(\theta), a_{t-1}(\theta))$ is bounded, a large change in x_1 will produce a large change in all the residuals $u_t(\theta)$. In the AR(1) model a change in one observation will produce changes in only two residuals.

A robust estimate for the MA(1) model will be introduced in Section 5, based on residuals that depend on a finite dimensional marginal empirical distribution.

In Section 2 we propose a new metric π_{d_n} on X^n. We compare this metric with those used by Hampel (1971).

In Section 3 we give a general definition of qualitative robustness which covers the case of dependent observations. We show that in the i.i.d. case our definition of robustness using the metric π_{d_n} is equivalent to Hampel's π-robustness concept.

In Section 4 we propose a different approach to qualitative robustness based on the concept of resistance, see Mosteller and Tukey (1977). The basic idea is to require that the estimate change only by a small amount when the sample is changed by replacing a small fraction of the sample by arbitrary large outliers or by perturbing all the observations of the sample with small errors (round-off errors). We give two definitions which formalize the resistance concept in probabilistic terms, calling them weak and strong resistance. The advantage of this approach is that we may require as a condition for resistance that the estimator itself as well as its distribution be insensitive to outliers or round-off errors. Moreover, the two definitions of resistance are based on much more elementary mathematical concepts than those used in the definition of qualitative robustness and, therefore, they are more likely to be easily grasped by applied scientists. In this regard we note that Donoho (1982) and Donoho and Huber (1983) have recently stressed finite sample breakdown points of estimates for similar reasons (i.e., transparency and simplicity of concept).

We shall show that weak resistance, which may be considered a generalization of Condition B given by Hampel (1971), is equivalent to robustness based on the proposed metric π_{d_n} if the parameter space is a subset of R^k. It is also shown that strong and weak resistance are equivalent in an i.i.d. model. We conjecture that this equivalence should hold even for more general stationary and ergodic processes. Asymptotic versions of both qualitative resistance definitions are given and these will be usually easier to verify for estimates implicitly defined by a system of equations.

In Section 5 it is noted that the continuity condition given in Papantoni-Kazakos and Gray (1979) is sufficient for asymptotic resistance. There we also establish asymptotic strong robustness for a large class of estimates which includes the GM-estimates used by Denby and Martin (1979), Martin (1982) and Bustos (1982). We also state a similar result for a class of estimates which are defined similarly to the class of scale estimates for regression proposed by Rousseeuw and Yohai (1984).

2. Distances between probabilities. Let X be the sample space and d be a distance on X. We shall assume throughout this paper that (X, d) is a complete and separable metric space (polish space). Let X^n and X^∞ be the cartesian product of copies of X, respectively. $\mathscr{F}$ will denote the Borel σ-field on X and $\mathscr{F}^n$ and $\mathscr{F}^\infty$ the corresponding product σ-fields on X^n and X^∞. For any measurable space $(\Omega, \mathscr{A})$, let $\mathscr{P}(\Omega)$ be the class of all probabilities on $\mathscr{A}$. If μ and ν are in $\mathscr{P}(\Omega)$, $\mathscr{P}(\mu, \nu)$ denotes the class of all the probabilities P on $(\Omega \times \Omega, \mathscr{A} \times \mathscr{A})$ with marginals μ and ν.

If (X, d) is a metric space, then the Prohorov distance π_d between μ and ν, where μ and $\nu \in \mathscr{P}(X)$, is defined by

$$\pi_d(\mu, \nu) = \inf\{\varepsilon: \mu(A) \le \nu(\mathscr{V}(A, \varepsilon, d)) + \varepsilon \quad \forall A \in \mathscr{F}\},$$

where $\mathscr{V}(A, \varepsilon, d) = \{x \in X: d(x, A) < \varepsilon\}$.

Strassen (1965) establishes that if (X, d) is a polish space, then π_d is given by

$$\pi_d(\nu, \mu) = \inf\{\varepsilon: \exists\, P \in \mathscr{P}(\mu, \nu) \text{ satisfying } P(\{(x, x'): d(x, x') \ge \varepsilon\}) \le \varepsilon\}.$$

Given $x^n = (x_1, \dots, x_n) \in X^n$ and $k \le n$, the kth empirical marginal distribution induced by x^n is denoted by $\mu_k[x^n]$ and is defined as the element of $\mathscr{P}(X^n)$ which assigns mass $1/(n - k + 1)$ to each sample $(x_{j+1}, x_{j+2}, \dots, x_{j+k})$, $0 \le j \le n - k$.

Metrics on X^n and $\tilde{X}^n$. Given (X, d) we will consider the following metric on X^n:

$$d_n(x^n, y^n) = \inf\{\varepsilon: \#\{i: d(x_i, y_i) \ge \varepsilon\}/n \le \varepsilon\}. \tag{2.1}$$

Let now $\tilde{X}^n$ be the space X^n modulo the permutation of coordinates. Hampel (1971) defines the following distance on $\tilde{X}^n$, which we denote by $\tilde{d}_n$:

$$\tilde{d}_n(\tilde{x}^n, \tilde{y}^n) = \pi_d(\mu_1[x^n], \mu_1[y^n]). \tag{2.2}$$

REMARK 2.1. Two points of X^n are close in the metric d_n if all the coordinates except a small fraction are close. Therefore, this notion of closeness corresponds to the type of errors which are considered in the intuitive notion of robustness or resistance. The following lemma gives the relationship between d_n and $\tilde{d}_n$.

LEMMA 2.1. *Let $\mathscr{P}_n$ be the set of all permutations of the first n positive integers. Given x^n and y^n in X^n, if p is in $\mathscr{P}_n$, we denote $y_p^n = (y_{p(1)}, \dots, y_{p(n)})$. Then we have*

$$\tilde{d}_n(x^n, y^n) = \min_{p \in \mathscr{P}_n} d_n(x^n, y_p^n). \tag{2.3}$$

PROOF. It is enough to show that for any $\delta > 0$,

$$d_n(x^n, y^n) \le \delta \Rightarrow \pi_d(\mu_1[x^n], \mu_1[y^n]) \le \delta \tag{2.4}$$

and

$$\pi_d(\mu_1[x^n], \mu_1[y^n]) < \delta \Rightarrow \exists\, p \in \mathscr{P}_n\colon d_n(x^n, y_p^n) < \delta. \tag{2.5}$$

Suppose that $d_n(x^n, y^n) \le \delta$. Then if $S = \{i\colon d(x_i, y_i) < \delta\}$ we have that $\#S/n > 1 - \delta$. Let R be the distribution on $X \times X$ which assigns probability $1/n$ to each pair (x_i, y_i), $1 \le i \le n$. Then $R \in \mathscr{P}(\mu_1[x^n], \mu_1[y^n])$. We also have that $R(d(x, y) < \delta) = \#S/n > 1 - \delta$. Then, by Strassen's theorem (1965), $\pi_d(\mu_1[x^n], \mu_1[y^n]) \le \delta$ and (2.4) holds.

For any p in $\mathscr{P}_n$, define $h(p) = \#\{i\colon d(x_i, y_{p(i)}) < \delta\}$ and $t = \max_{p \in \mathscr{P}_n} h(p)$.

We have to show that $t > n(1 - \delta)$ whenever $\pi_d(\mu_1[x^n], \mu_1[y^n]) < \delta$. Consider the map $f\colon \{1, 2, \dots, 2n\} \to X$ defined by $f(i) = x_i$ if $i \le n$, $f(i) = y_{i-n}$ if $i > n$ and put $\tilde{d}(i, j) = d(f(i), f(j))$.

By the definition of the Prohorov distance, we have

$$\pi_d(\mu_1[x^n], \mu_1[y^n]) = \pi_{\tilde{d}}(\nu_1, \nu_2), \tag{2.6}$$

where ν_1 and ν_2 are the empirical distributions induced by $(1, \dots, n)$ and $(n + 1, \dots, 2n)$, respectively. From Strassen's theorem (1965), it follows that the right-hand side of (2.6) is equal to $\inf\{\varepsilon$: there exists a doubly stochastic matrix (p_{ij}) such that $\sum p_{ij} I_{\{d(x_i, y_j) \ge \varepsilon\}} \le n\varepsilon\}$. Since by Birkhoff's theorem (1946) every doubly stochastic matrix is a convex combination of permutation matrices, the lemma follows. □

3. Generalization of Hampel's definition of robustness. Let $T_n\colon X^n \to \Lambda$ for $n \ge n_0$ be a sequence of estimators taking values in a polish space (Λ, λ). Given $\mu_n \in \mathscr{P}(X^n)$, we denote by $\mathscr{L}(T_n, \mu_n)$ the distribution of T_n under μ_n.

The following definition given by Bustos (1981) generalizes Hampel's concept of π-robustness.

DEFINITION. Let $\mu \in \mathscr{P}(X^\infty)$ and let ρ_n be a pseudometric on $\mathscr{P}(X^n)$ for all $n \ge n_0$. Then the sequence $(T_n)_{n \ge n_0}$ is *ρ_n-robust at μ* if given $\varepsilon > 0$, there exists

$\delta > 0$ such that

$$\nu_n \in \mathscr{P}(X^n) \wedge n \geq n_0 \wedge \rho_n(\mu_n, \nu_n) < \delta \Rightarrow \pi_\lambda(\mathscr{L}(T_n, \mu_n), \mathscr{L}(T_n, \nu_n)) \leq \varepsilon,$$

where μ_n is the nth order marginal of μ.

In the case of μ an i.i.d. process and $(T_n)_{n \geq n_0}$ invariant under permutation of coordinates, we get Hampel's definition taking as ρ_n the pseudometric $\bar{\rho}_n$ defined by $\bar{\rho}_n(\mu_n, \nu_n) = \pi_{\tilde{d}_n}(\tilde{\mu}_n, \tilde{\nu}_n)$, where $\tilde{\mu}_n$ and $\tilde{\nu}_n$ are the probabilities induced by μ_n and ν_n on $\tilde{X}_n$.

REMARK 3.1. Papantoni-Kazakos and Gray (1981) and Cox (1981) gave definitions which are generalizations of Hampel's concept of robustness instead of π-robustness. We consider that these definitions are not adequate for dependent observations (where the estimate may not be invariant under permutations of coordinates) since according to Papantoni-Kazakos and Gray and Cox, estimates which depend on a fixed finite set of coordinates may be robust and this contradicts the intuitive notion of robustness: A small proportion of observations should not affect the estimator too much [consider, for example, T_n: $X^n \to X$ defined by $T_n(x_1, \ldots, x_n) = x_1$]. In Boente, Fraiman and Yohai (1982) may be found a discussion on why the relevant concept of Hampel to be generalized for dependent observations is π-robustness and not robustness.

The following theorem shows that π_{d_n}-robustness is a natural generalization of Hampel's definition.

THEOREM 3.1. *Let μ be an i.i.d. process and $(T_n)_{n \geq 1}$ a sequence of estimates invariant by permutations of the coordinates. Then Hampel's definition of π-robustness ($\bar{\rho}_n$-robustness) is equivalent to π_{d_n}-robustness.*

PROOF. By Lemma 2.1 we have that $\pi_{d_n}(\mu_n, \nu_n) \leq \delta$ implies $\bar{\rho}_n(\mu_n, \nu_n) \leq \delta$. Thus, $\bar{\rho}_n$-robustness implies π_{d_n}-robustness.

Assume now that $(T_n)_{n \geq n_0}$ is π_{d_n}-robust at μ. We have to show that given $\varepsilon > 0$, there exists $\delta > 0$ such that $\forall\ n \geq n_0$, we have

$$(3.1)\quad \nu_n \in \mathscr{P}(X^n) \quad \text{and} \quad \pi_{\tilde{d}_n}(\tilde{\mu}_n, \tilde{\nu}_n) \leq \delta \Rightarrow \pi_\lambda(\mathscr{L}(T_n, \mu), \mathscr{L}(T_n, \nu_n)) \leq \varepsilon.$$

Since $(T_n)_{n \geq n_0}$ is π_{d_n}-robust we can find $\delta_0 > 0$ such that $\forall\ n \geq n_0$,

$$(3.2)\quad \pi_{d_n}(\mu_n, \nu_n) \leq \delta_0 \Rightarrow \pi_\lambda(\mathscr{L}(T_n, \mu), \mathscr{L}(T_n, \nu_n)) \leq \varepsilon.$$

We will show that (3.1) holds with $\delta = \delta_0$. Let ν_n be such that $\pi_{\tilde{d}_n}(\tilde{\mu}_n, \tilde{\nu}_n) \leq \delta_0$. Then, by Strassen's theorem, there exists $\tilde{R}_n \in \mathscr{P}(\tilde{\mu}_n, \tilde{\nu}_n)$ such that

$$(3.3)\quad \tilde{R}_n(\{(\tilde{x}^n, \tilde{y}^n) \in \tilde{X}^n \times \tilde{X}^n: \tilde{d}_n(\tilde{x}^n, \tilde{y}^n) \leq \delta_0\}) \geq 1 - \delta_0.$$

Let S_n: $X^n \to \tilde{X}^n$ be the canonical projection. By Lemma 2.1 we can define a function $g_n = (g_{n,1}, g_{n,2})$, g_n: $\tilde{X}^n \times \tilde{X}^n \to X^n \times X^n$, such that

(a) $S_n(g_{n,1}(\tilde{x}^n, \tilde{y}^n)) = \tilde{x}^n$,

(b) $S_n(g_{n,2}(\tilde{x}^n, \tilde{y}^n)) = \tilde{y}^n$,

(c) $\tilde{d}_n(\tilde{x}^n, \tilde{y}^n) = d_n(g_{n,1}(\tilde{x}^n, \tilde{y}^n), g_{n,2}(\tilde{x}^n, \tilde{y}^n))$.

It is easy to show that g_n may be chosen to be measurable. Let R_n be the probability induced by g_n on $X^n \times X^n$ when $\tilde{X}^n \times \tilde{X}^n$ is endowed with the probability $\tilde{R}_n$. Then as μ is an i.i.d. process, (a) and (b) entail that the first marginal of R_n is μ_n. If ν_n^* denotes the second marginal of R_n we have $\tilde{\nu}_n^* = \tilde{\nu}_n$. Finally, (c) and (3.3) entail

$$R_n(\{(x^n, y^n) \in X^n \times X^n: d_n(x^n, y^n) \le \delta_0\}) \ge 1 - \delta_0. \tag{3.4}$$

Using Strassen's theorem we have $\pi_{d_n}(\mu_n, \nu_n^*) \le \delta_0$ and, therefore, by (3.2), $\pi_\lambda(\mathscr{L}(T_n, \mu), \mathscr{L}(T_n, \nu_n^*)) \le \varepsilon$. Since T_n is invariant by permutation of coordinates $\mathscr{L}(T_n, \nu_n^*) = \mathscr{L}(T_n, \nu_n)$ and then (3.1) holds. □

REMARK 3.2. Papantoni-Kazakos and Gray (1979) define a concept of robustness using the Vassershtein distance $\rho_{d,n}^*$ on $\mathscr{P}(X^n)$. It is possible to show [see Boente, Fraiman and Yohai (1982)] that π_{d_n}-robustness implies $\rho_{d,n}^*$-robustness and if d is bounded, both concepts are equivalent. A shortcoming of the notion of robustness based on $\rho_{d,n}^*$ is the lack of invariance with respect to equivalent d metrics.

4. Qualitative resistance. Here, we propose a different approach to qualitative robustness, based on the concept of resistance [Tukey (1976)] which seems to capture better its intuitive meaning. Instead of considering the insensitivity of the estimates with respect to small changes in the distribution of the process, we look at how insensitive they are at a given sample point $x \in X^\infty$, when

(a) all the observations have small changes,
(b) a small fraction of observations suffer large changes.

Consider $x^n \in X^n$ and let $\mathscr{V}(x^n, \delta, d_n)$ be the open sphere of center x^n and radius δ corresponding to the metric d_n. Define

$$S_n(\delta, x^n) = \sup\{\lambda(T_n(y^n), T_n(z^n)): y^n, z^n \in \mathscr{V}(x^n, \delta, d_n)\}.$$

According to Lemma 4.1(i), $S_n(\delta, \cdot)$ is lower semicontinuous and, therefore, measurable.

DEFINITION 1. Let $x \in X^\infty$ and $x^{(n)}$ be the first n coordinates of x. Then $(T_n)_{n \ge n_0}$ is *resistant* at x if given $\varepsilon > 0$, there exists $\delta > 0$ such that $S_n(\delta, x^{(n)}) \le \varepsilon$, $\forall\, n \ge n_0$.

We will now give two definitions which formalize the concept of resistance at a given probability $\mu \in \mathscr{P}(X^\infty)$.

DEFINITION 2. Let $\mu \in \mathscr{P}(X^\infty)$; then $(T_n)_{n \ge n_0}$ is *strongly resistant* at μ if $\mu(\{x \in X^\infty: T_n \text{ is resistant at } x\}) = 1$.

DEFINITION 3. Let $\mu \in \mathscr{P}(X^\infty)$; then $(T_n)_{n \ge n_0}$ is *weakly resistant* at μ if given $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\mu_n(\{x^n \in X^n: S_n(\delta, x^n) \le \varepsilon\}) \ge 1 - \varepsilon \qquad \forall\, n \ge n_0.$$

REMARK 4.1. According to what has been seen in Section 1, the M-estimates with convex ρ for the AR(1) model are not weakly or strongly resistant. The same occurs with the GM-estimates for the MA(1) model with $\eta(u, v)$ monotone in both variables.

The advantage of these definitions is that they require that the estimator itself be insensitive to errors of type (a) and (b) mentioned earlier, while the other definitions only require insensitivity of the law of the estimators. However, we will show later the equivalence of weak resistance with π_{d_n}-robustness. We also show that at least in the i.i.d. case strong and weak resistance are equivalent.

The following elementary proposition, which we give without proof [see Boente, Fraiman and Yohai (1982)], gives a characterization of strong resistance.

PROPOSITION 4.1. (a) $(T_n)_{n \geq n_0}$ *is strongly resistant at* μ *if given* $\varepsilon > 0$, *there exists* $\delta > 0$ *such that*

$$\mu\left(\bigcap_{n \geq n_0} \{x \in X^\infty \colon S_n(\delta, x^{(n)}) \leq \varepsilon\} \right) \geq 1 - \varepsilon. \tag{4.1}$$

(b) *Then strong resistance of* $(T_n)_{n \geq n_0}$ *at* μ *implies weak resistance at* μ.

Similar to Papantoni-Kazakos and Gray (1979), we give asymptotic versions of Definitions 1, 2 and 3.

DEFINITION 4. Let $x \in X^\infty$; then $(T_n)_{n \geq n_0}$ *is asymptotically resistant* (*AR*) *at* x if given $\varepsilon > 0$, there exist $\delta > 0$ and $n_1 = n_1(\varepsilon, x)$ such that

$$S_n(\delta, x^{(n)}) \leq \varepsilon \qquad \forall\, n \geq n_1.$$

DEFINITION 5. Let $\mu \in \mathscr{P}(X^\infty)$; then $(T_n)_{n \geq n_0}$ is *asymptotically strongly resistant* (*ASR*) *at* μ if

$$\mu(\{x \in X^\infty \colon T_n \text{ is AR at } x\}) = 1.$$

DEFINITION 6. Let $\mu \in \mathscr{P}(X^\infty)$; then $(T_n)_{n \geq n_0}$ is *asymptotically weakly resistant* (*AWR*) *at* μ if given $\varepsilon > 0$, there exist $\delta > 0$ and $n_1 = n_1(\varepsilon)$ such that

$$\mu_n(\{x^n \in X^n \colon S_n(\delta, x^n) \leq \varepsilon\}) \geq 1 - \varepsilon \qquad \forall\, n \geq n_1.$$

As a direct consequence of these definitions we have the following proposition that we give without proof and which extends analogous results obtained by Hampel (1971) and Papantoni-Kazakos and Gray (1979).

PROPOSITION 4.2. $(T_n)_{n \geq n_0}$ *is strongly* (*weakly*) *resistant at* μ *if and only if*

(i) T_n *is asymptotically strong* (*weak*) *resistant at* μ;
(ii) T_n *is a continuous function of* x^n, *for each* n.

The following theorem shows that for the i.i.d. case, strong and weak resistance definitions are equivalent.

THEOREM 4.1. *Let $\mu \in \mathscr{P}(X^\infty)$ correspond to an i.i.d. process and let $(T_n)_{n \geq n_0}$ be invariant under permutations of coordinates and weakly resistant at μ. Then $(T_n)_{n \geq n_0}$ is also strongly resistant at μ.*

PROOF. Let $\varepsilon > 0$. We are going to show that there exists $\delta > 0$ such that (4.1) holds. Since $(T_n)_{n \geq n_0}$ is weakly resistant, we can find δ^* such that

$$\mu_n(\{x^n: S_n(\delta^*, x^n) \leq \varepsilon\}) \geq 1 - \varepsilon. \tag{4.2}$$

We will show by applying a suitable form of the law of large numbers that an even stronger inequality holds, namely,

$$\mu_n(\{x^n: S_n(\delta^*/2, x^n) \leq \varepsilon\}) \geq 1 - ab^n \qquad \forall\, n \geq n_1, \tag{4.3}$$

for suitable $a > 0$, $0 < b < 1$ and n_1. Then, let

$$B_m = X^\infty - \{x: S_n(\delta^*/2, x^{(n)}) \leq \varepsilon\ \forall\, n \geq m\}.$$

Since

$$\mu(B_m) \leq \sum_{n \geq m} \mu_n(\{x^n: S_n(\delta^*/2, x^n) > \varepsilon\}) \leq ab^m/(1 - b),$$

there exists $n_2 \geq \max(n_1, n_0)$ such that

$$\mu\left(\bigcap_{n \geq n_2} \{x: S_n(\delta^*/2, x^{(n)}) \leq \varepsilon\}\right) \geq 1 - \varepsilon/2. \tag{4.4}$$

Finally, we can find $\delta \geq \delta^*/2$ such that

$$\mu(\{x: S_n(\delta, x^{(n)}) \leq \varepsilon\}) \geq 1 - \varepsilon/(2(n_2 - n_0)) \qquad \forall\, n_0 \leq n \leq n_2. \tag{4.5}$$

Then (4.1) may be derived from (4.4) and (4.5).

In order to prove (4.3), note that since X is a polish space, there exists a compact K such that

$$\mu_1(K) \geq 1 - \delta/8. \tag{4.6}$$

We can find a partition $K_1, \dots, K_h$ of K such that each K_i, $1 \leq i \leq h$, has diameter $\leq \delta^*/2$. Let K_0 be the complement of K and $m_i = \mu_1(K_i)$, $0 \leq i \leq h$.

Given $x^n = (x_1, \dots, x_n)$, define $M_{in}(x^n) = \sum_{j=1}^n I_{K_i}(x_j)/n$, $i = 0, 1, \dots, h$. By a well known form of the strong law of the large numbers for Bernoulli variables, there exist $a > 0$ and $0 < b < 1$ such that if $R_n = \bigcap_{i=0}^h \{x^n: |M_{in}(x^n) - m_i| \leq \delta^*/(8h)\}$, then

$$\mu_n(R_n) \geq 1 - ab^n \qquad \forall\, n. \tag{4.7}$$

Now, look at a fixed $x^{*n} \in R_n$ and let P_n be the set of all points obtained by permutation of the coordinates of x^{*n}. We will show that

$$R_n \subset \mathscr{V}(P_n, \delta^*/2, d_n). \tag{4.8}$$

Let $y^n \in R_n$; then $|M_{in}(x^{*n}) - M_{in}(y^n)| \le \delta^*/(4h)$. Define $Q_i = \{j: y_j \in K_i\}$, $0 \le i \le h$. Then there exist sets Q_i^*, $0 \le i \le h$, such that

$$\#Q_i^* = nM_{in}(x^{*n}), \qquad 0 \le i \le h, \tag{4.9}$$

and

$$\#(Q_i^* \vartriangle Q_i) \le n\delta^*/(4h). \tag{4.10}$$

By (4.9) there exists a point $\bar{x}^n = (\bar{x}_1, \dots, \bar{x}_n) \in P_n$, such that if $j \in Q_i^*$, then $\bar{x}_j \in K_i$, $0 \le i \le h$. Since the diameter of K_i, $1 \le i \le h$, is smaller than $\delta^*/2$, we have that $\{i: |\bar{x}_i - y_i| \ge \delta^*/2\} \subset (\bigcup_{i=1}^h (Q_i^* \vartriangle Q_i)) \cup Q_0^*$. Then (4.6), (4.9) and (4.10) imply that

$$\#\{i: |\bar{x}_i - y_i| \ge \delta^*/2\} \le n\delta^*/4 + \#Q_0^* \le n\delta^*/4 + n\delta^*/4 \le n\delta^*/2.$$

Therefore, $d_n(\bar{x}^n, y^n) \le \delta^*/2$ and (4.8) is true.

From (4.2) and (4.7) we can find $n_1 \ge n_0$ and x^{*n} in $R_n \cap \{S_n(\delta^*, x^n) \le \varepsilon\}$ $\forall\, n \ge n_1$. Since $\{S_n(\delta^*, x^n) \le \varepsilon\}$ are invariant under permutation of coordinates, we have

$$P_n \subset \{S_n(\delta^*, x^n) \le \varepsilon\}. \tag{4.11}$$

On the other hand, it is clear that

$$\{x^n: S_n(\delta^*/2, x^n) \le \varepsilon\} \supset \mathscr{V}(\{x^n: S_n(\delta^*, x^n) \le \varepsilon\}, \delta^*/2, d_n). \tag{4.12}$$

Combining the last two equations with (4.7) and (4.8) we get (4.3). □

REMARK 4.2. Proposition 4.1(b) and Theorem 4.1 also hold if strong and weak resistance are replaced by asymptotically strong and weak resistance.

THEOREM 4.2. *Let* $\mu \in \mathscr{P}(X^\infty)$.

(i) *If* $(T_n)_{n \ge n_0}$ *is weakly resistant at* μ, *then it is* π_{d_n}*-robust at* μ.

(ii) *If* $(T_n)_{n \ge n_0}$ *is asymptotically* π_{d_n}*-robust and consistent at* μ, *then it is also asymptotically weakly resistant at* μ.

To prove Theorem 4.2 we need the following lemma.

LEMMA 4.1. *Let* (A, ρ) *and* (Λ, λ) *be two polish spaces and* $T: A \to \Lambda$ *a measurable function with respect to the Borel* σ*-field. For any* $a \in A$, $\delta > 0$ *let* $S_\delta(a) = \sup\{\lambda(T(b), T(c)): b, c \in \mathscr{V}(a, \delta, \rho)\}$. *Then we have*:

(i) S_δ *is lower semicontinuous and, therefore, measurable.*

(ii) *For any* $\delta > 0$ *there exist measurable functions* U_j: $A \to A$, $j = 1, 2$, *such that for all* a (a) $U_j(a) \in \mathscr{V}(a, 2\delta, \rho)$, $j = 1, 2$, *and* (b) $\lambda(T(U_1(a)), T(U_2(a))) \ge S_\delta(a)/2$.

The proof may be found in Boente, Fraiman and Yohai (1982).

PROOF OF THEOREM 4.2. (i) Given $\varepsilon > 0$, choose $\delta_1 > 0$ verifying $\mu_n(B) > 1 - \varepsilon/2$ with $B = \{x^n: S_n(\delta_1, x^n) \le \varepsilon\}$ and let $\delta = \min(\delta_1, \varepsilon/2)$. By Strassen's theorem we have that for any $\nu_n \in \mathscr{P}(X^n)$ such that $\pi_{d_n}(\mu_n, \nu_n) < \delta$, there exists

$R \in \mathscr{P}(\mu_n, \nu_n)$ such that $R(\Delta) \geq 1 - \delta$, where $\Delta = \{(x^n, y^n): d_n(x^n, y^n) \leq \delta\}$. Therefore, we have $R(\Delta \cap (B \times X^n)) \geq 1 - \varepsilon$, which implies

$$\pi_\lambda(\mathscr{L}(T_n, \mu_n), \mathscr{L}(T_n, \nu_n)) \leq \varepsilon.$$

(ii) Since $(T_n)_{n \geq n_0}$ is π_{d_n}-robust at μ, given $\varepsilon > 0$, there exists $\delta_0 > 0$ such that

$$\pi_{d_n}(\mu_n, \nu_n) < \delta_0 \Rightarrow \pi_\lambda(\mathscr{L}(T_n, \mu_n), \mathscr{L}(T_n, \nu_n)) < \varepsilon/4.$$

By Lemma 4.1 (ii) we can find for $\delta = \delta_0/2$ and for any $n \geq n_0$ a pair of measurable functions U_n, W_n: $X^n \to X^n$, such that

$$d_n(U_n(x^n), x^n) \leq 2\delta, \qquad d_n(W_n(x^n), x^n) \leq 2\delta \tag{4.13}$$

and

$$\lambda(T_n^-(x^n), T_n^+(x^n)) \geq S_n(\delta, x^n)/2,$$

where $T_n^- = T_n \circ U_n$ and $T_n^+ = T_n \circ W_n$.

Denote $\mu_n^- = \mathscr{L}(U_n(x^n), \mu_n)$ and $\mu_n^+ = \mathscr{L}(W_n(x^n), \mu_n)$. By (4.13) we have $\pi_{d_n}(\mu_n^-, \mu_n) < \delta_0$ and $\pi_{d_n}(\mu_n^+, \mu_n) < \delta_0$, which implies

$$\pi_\lambda(\mathscr{L}(T_n, \mu_n), \mathscr{L}(T_n^+, \mu_n)) = \pi_\lambda(\mathscr{L}(T_n, \mu_n), \mathscr{L}(T_n, \mu_n^+)) < \varepsilon/4$$

and

$$\pi_\lambda(\mathscr{L}(T_n, \mu_n), \mathscr{L}(T_n^-, \mu_n)) = \pi_\lambda(\mathscr{L}(T_n, \mu_n), \mathscr{L}(T_n, \mu_n^-)) < \varepsilon/4.$$

Finally, the consistency of T_n to T_0 at μ implies that there exists $n_1 \geq n_0$ such that

$$\pi_\lambda(\mathscr{L}(T_n^+, \mu_n), \delta_{T_0}) \leq \varepsilon/2 \quad \text{and} \quad \pi_\lambda(\mathscr{L}(T_n^-, \mu_n), \delta_{T_0}) \leq \varepsilon/2 \qquad \forall\, n \geq n_1,$$

where δ_{T_0} is the one point mass at T_0. Therefore, $\mu\{\lambda(T_n^-, T_n^+) > \varepsilon\} < \varepsilon$ $\forall\, n \geq n_1$. □

5. Applications. A standard argument shows that the generalization of Hampel's (1971) continuity condition given by Papantoni-Kazakos and Gray (1979), Cox (1981) and Bustos (1981) implies ASR for stationary and ergodic processes [see Boente, Fraiman and Yohai (1982)]. However, in this section, we will prove the ASR property directly from the definition. We will give sufficient conditions for the ASR of a class of estimates, called GM-estimates, which include the proposals given by Denby and Martin (1979), Martin (1982) and Bustos (1982) for autoregressive models and by Krasker (1980), Maronna and Yohai (1981) and Krasker and Welsch (1982) for regression models.

From now on (X, d) will be an Euclidean space, which we assume to be either (a) $X = R^{p+1}$ or (b) $X = R$ when we are interested in estimating the parameters of a linear model or an autoregressive model of order p, respectively, from a sample obtained from a distribution in the neighborhood of the ideal model.

The parameters to be estimated will be $\alpha = (\theta, \sigma)$, where $\theta \in \Theta$ is an open set in R^p and $\sigma > 0$ is a scale parameter for the errors. Then $\Lambda = \Theta \times R^+$ endowed with the Euclidean distance.

Let us define the scaled residual map as the function

$$r: R^p \times R \times R^p \times R^+ \to R, \qquad r(z, y, \theta, \sigma) = (y - \theta' z)/\sigma.$$

Let ϕ: $R^p \times R \to R$, χ: $[0, +\infty) \to R$ be real functions and φ: $R^{p+1} \times \Lambda \to R^{p+1}$ be such that

$$\varphi(w, \alpha) = (\phi(Az, r(z, y, \theta, \sigma))Az, \chi(r(z, y, \theta, \sigma)))',$$

where $w = (y, z')'$, $z \in R^p$, $y \in R$ and A is a nonsingular matrix.

Let $x_t^h = (x_t, \dots, x_{t-h})'$ for $h \leq t$ and k be a fixed integer which will be 0 or p for cases (a) or (b), respectively.

For $x^n \in X^n$ and $\alpha \in \Lambda$ we define

$$g_n(x^n, \alpha) = \sum_{t=k+1}^{n} \varphi(x_t^k, \alpha)/(n-k).$$

The GM-estimate of α is defined as a solution of

$$g_n(x^n, T_n) = 0. \tag{5.1}$$

More generally, if there is no solution of (5.1) we will consider any sequence of estimates $(T_n)_{n \geq k+1}$ such that

$$\lim_n \sup_{x^n \in X^n} | |g_n(x^n, T_n)| - i_n(x^n)| = 0, \tag{5.2}$$

where $i_n(x^n) = \inf_{\alpha \in \Lambda} |g_n(x^n, \alpha)|$.

The matrix $\Sigma = AA'$ is a well chosen scatter matrix of the vector Z_t, where Z_t corresponds to the last p coordinates of the vector x_t^k, i.e., a scatter matrix of the regression or autoregression variables, respectively. Krasker (1980), Krasker and Welsch (1982), Ronchetti and Rousseeuw (1982) and Samarov (1983) give optimal A matrices and ϕ functions for the linear model under different criteria.

In the case of autoregression [see Künsch (1984)], the matrix A depends on α. For example, in a Gaussian AR(1) process, $A(\alpha) = (1 - \theta^2)^{1/2}/\sigma$. However, in this case the equation $E_\mu(\varphi(x_t^k, \alpha)) = 0$ will typically have an additional degenerate solution, e.g., in the AR(1) case $\theta^2 = 1$. Therefore, the consistency and resistance of the estimates will depend also on how the estimate is chosen among the set of solutions of (5.1). This degenerate solution may be avoided by multiplying the function φ by the matrix $\begin{pmatrix} A^{-1} & 0 \\ 0 & 1 \end{pmatrix}$. In this case an analogous statement to Theorem 5.1 should be obtained.

Another possible solution will be to replace A by any consistent and ASR sequence A_n. For the sake of simplicity, in this section we will study only the case of a fixed A matrix.

If we replace A by a sequence A_n the results in this section still hold provided the function ϕ verifies condition E2 of Bustos (1982) instead of E2(ii) which follows.

We will need the following assumptions:

E1. For each $z \in R^p$, $u \to \phi(z, u)$ is odd, uniformly continuous and $\phi(z, u) \geq 0$ for $u \geq 0$.

E2. (i) $(z, u) \to \phi(z, u)$ is bounded.

(ii) There exists a constant K^* such that

$$|\phi(z_1, u)z_1 - \phi(z_2, u)z_2| \leq K^*|z_1 - z_2|/\min(|z_1|, |z_2|),$$
$$|\phi(z, u_1)z - \phi(z, u_2)z| \leq K^*|u_1 - u_2|/\min(|u_1|, |u_2|).$$

E3. For each z, $u \to \phi(z, u)/u$ is nonincreasing and there exists u_0 such that $\phi(z, u_0)/u_0 > 0$ for all z.

E4. χ is bounded, continuous, even and increasing on $\{x\colon -c_1 < \chi(x) < c_2\}$, where $c_2 = \sup \chi(x)$ and $-c_1 = \chi(0)$, $0 < c_1, c_2$.

χ is differentiable with $x \to x\chi'(x)$ continuous and bounded. Also, $\chi(|u_0|) > 0$.

H1. $x_1, \ldots, x_n, \ldots$ is a stationary and ergodic stochastic process with probability measure $\mu \in \mathscr{P}(X^\infty)$.

H2. $\mu(a'z_t = 0) = 0$ for all $a \in R^p$.

H3. $\mu((1, -a')x_t^k = 0) = 0$.

H4. There exists a unique α_0 such that $g(\alpha_0) = E_\mu \varphi(x_t^k, \alpha_0) = 0$.

As noted by Bustos (1982), Theorem 5 of Maronna and Yohai (1981) shows that E1–E4 are sufficient conditions for H4. Moreover, if the observations are from the ideal model and the residuals have a symmetric distribution, H4 holds for the true value of the parameters.

Two examples of such GM-estimates are:

(i) Mallows type estimates are defined by $\phi(y, u) = \psi_1(u)\psi_2(|y|)/|y|$, where $\psi_i\colon R \to R$, $i = 1, 2$. In order to obtain E1–E4 it is sufficient to assume $\psi_i(u) \geq 0$ for $u \geq 0$; ψ_i is odd, bounded, continuously differentiable with $u\psi_i'(u) \leq \psi_i(u)$ for $i = 1, 2$; also, ψ_1 is uniformly continuous and there exists u_0 such that $\psi_1(u_0) > 0$.

(ii) Hampel–Krasker type estimates are defined by $\phi(y, u) = \psi(|y|u)/|y|$. In order to obtain E1–E4 it suffices to require that ψ be odd, bounded and nondecreasing in $[0, +\infty)$ with $\psi(u_0) > 0$, uniformly continuous, continuously differentiable and $u\psi'(u) \leq \psi(u)$ for all u.

The following proposition shows that the function $g_n(x^n, \alpha)$ is uniformly equicontinuous with respect to the metric d_n.

PROPOSITION 5.1. *Assume* E2, E4 *and* H1–H3. *Then there exist* $N_1 \subset X^\infty$ *and* $\mu(N_1) = 0$, *verifying that for any* $x \notin N_1$ *and* $\varepsilon > 0$, *there exist* $\delta > 0$ *and* n_0 *such that for* $n \geq n_0$, $y^n \in X^n$ *and* $d_n(x^{(n)}, y^n) \leq \delta$ *implies*

$$|g_n(x^{(n)}, \alpha) - g_n(y^n, \alpha)| < \varepsilon, \qquad \forall\, \alpha \in \Lambda.$$

The proof may be found in the Appendix.

REMARK 5.1. In Bustos (1982) for autoregression and in Maronna and Yohai (1981) for regression, it is shown that under E2–E4 and H1–H3 there exists a compact set $K \subset \Lambda$ such that

$$\liminf_{n \to \infty} \inf_{\alpha \notin K} |g_n(x^{(n)}, \alpha)| > 0 \quad \text{a.s. } (\mu). \tag{5.3}$$

THEOREM 5.1. *Suppose that* $(T_n)_{n>k}$ *satisfies* (5.2); *then* H1–H4 *together with* (5.3) *and the conclusion of Proposition* 5.1 *imply that* $(T_n)_{n>k}$ *is ASR at* μ.

PROOF. The ergodic theorem implies

$$\lim_{n\to\infty} g_n\big(x^{(n)}, \alpha_0\big) = 0 \quad \text{a.s.} \tag{5.4}$$

Therefore, by (5.2) we have

$$\lim_{n\to\infty} g_n\big(x^{(n)}, T_n(x^{(n)})\big) = 0 \quad \text{a.s.} \tag{5.5}$$

As in the proof of Theorem 2 of Huber (1967) but using the ergodic theorem instead of the law of the large numbers it may be proved that for all $\varepsilon > 0$ and each compact set $K \subset \Lambda$,

$$\liminf_{n\to\infty} \inf_{(|\alpha-\alpha_0|\geq\varepsilon)\cap K} |g_n(x^{(n)}, \alpha)| > 0 \quad \text{a.s.},$$

which implies

$$\liminf_{n\to\infty} \inf_{|\alpha-\alpha_0|\geq\varepsilon} |g_n(x^{(n)}, \alpha)| > 0 \quad \text{a.s.} \tag{5.6}$$

Let $N_2^c \subset X^\infty$ be the set where (5.3)–(5.6) hold, and $N_3 = N_1 \cup N_2$. Given $x \notin N_3$ and $\varepsilon > 0$ by (5.4) and (5.6) there exist $n_0 \in N$ and $\eta > 0$ such that

$$\forall\, n \geq n_0, \qquad \inf_{|\alpha-\alpha_0|\geq\varepsilon/2} |g_n(x^{(n)}, \alpha)| > \eta, \qquad |g_n\big(x^{(n)}, \alpha_0\big)| < \eta/4. \tag{5.7}$$

Therefore, by (5.5) there exists n_1 such that

$$|T_n(x^{(n)}) - \alpha_0| < \varepsilon/2 \qquad \forall\, n \geq n_1. \tag{5.8}$$

As $x \notin N_1$, Proposition 5.1 implies that there exist $n_2 \geq \max(n_0, n_1)$ and $\delta > 0$ such that

$$\forall\, n \geq n_2, \qquad d_n\big(x^{(n)}, y^n\big) \leq \delta \Rightarrow \sup_{\alpha\in\Lambda} |g_n(x^{(n)}, \alpha) - g_n(y^n, \alpha)| < \eta/4. \tag{5.9}$$

From (5.7) and (5.9) we have that for all $n \geq n_2$, $d_n(x^{(n)}, y^n) < \delta$ implies that $\inf_{|\alpha-\alpha_0|\geq\varepsilon/2}|g_n(y^n, \alpha)| > 3\eta/4$ and $|g_n(y^n, \alpha_0)| < \eta/2$. Therefore, by (5.2) there exists $n_3 \geq n_2$ such that

$$\forall\, n \geq n_3, \qquad d_n\big(x^{(n)}, y^n\big) < \delta \Rightarrow |T_n(y^n) - \alpha_0| < \varepsilon/2,$$

which together with (5.8) implies the desired result. □

REMARK 5.2. H1–H4 together with (5.3) imply the strong consistency of $(T_n)_{n>k}$ as is shown by (5.8).

When we consider redescending ψ functions, H4 does not hold. In this case (5.1) will have solutions which are essentially different. In order to study the properties of the estimate, we should specify which solution we are considering. In this paper we will consider the solution of (5.1) which is closest to an initial consistent and ASR estimate. In Theorem 5.2 we show that these estimators are ASR when the parameter space $\Lambda \subset R$.

Let η: $R^2 \times \Lambda \to R$ be a real function. For $x^n \in R^n$ and $\alpha \in \Lambda$ define

$$g_n(x^n, \alpha) = \sum_{t=2}^{n} \eta(x_t^1, \alpha)/(n-k), \qquad g(\alpha) = E_\mu(\eta(x_t^1, \alpha)).$$

We will need the following assumption.

H5. (i) η is bounded, continuous and satisfies E2.
(ii) There exists $\alpha_0 \in \Lambda$ such that $g(\alpha_0) = 0$ and $g(\alpha)$ is strictly monotone in a neighborhood of α_0.

For each $\varepsilon > 0$ we define

$$N_\varepsilon^c = \{x \in R^\infty: \exists\, n_0, \delta > 0: n \geq n_0 \text{ and } d_n(x^{(n)}, y^n) \leq \delta \Rightarrow g_n(y^n, \alpha) = 0 \text{ for some } \alpha \in [\alpha_0 - \varepsilon, \alpha_0 + \varepsilon]\}.$$

LEMMA 5.1. *Assume* H1, H5 *and* H2 *and* H3 *with* $p = 1$. *Then* $\mu(N_\varepsilon) = 0$.

PROOF. Without loss of generality we may suppose that $\varepsilon < \varepsilon_0$ and that $g(\alpha)$ is strictly increasing in $(\alpha_0 - \varepsilon_0, \alpha_0 + \varepsilon_0)$. Then the ergodic theorem implies

$$\lim_{n\to\infty} g_n(x^{(n)}, \alpha_0 + \varepsilon) = g(\alpha_0 + \varepsilon) = a > 0 \quad \text{a.s.} \tag{5.10}$$

and

$$\lim_{n\to\infty} g_n(x^{(n)}, \alpha_0 - \varepsilon) = g(\alpha_0 - \varepsilon) = b < 0 \quad \text{a.s.} \tag{5.11}$$

Let N^c be the set where (5.10) and (5.11) hold and $M = N \cup N_1$ with N_1 defined in Proposition 5.1. If $x \notin M$, Proposition 5.1, (5.10) and (5.11) imply that there exist n_1 and $\delta > 0$ such that

$$\forall\, n \geq n_1, \qquad d_n(x^{(n)}, y^n) \leq \delta \Rightarrow \begin{cases} g_n(y^n, \alpha_0 + \varepsilon) \geq a/2, \\ g_n(y^n, \alpha_0 - \varepsilon) \leq b/2. \end{cases}$$

Since g_n is continuous, we have $x \in N^c$. □

Let $(T_n^*)_{n>k}$ be an initial sequence of strongly consistent estimates of α_0 which is ASR at μ and $A_n = \{x^n \in R^n: \exists\, \alpha: g_n(x^n, \alpha) = 0\}$. We define $T_n(x^n)$ as the solution of $g_n(x^n, \alpha) = 0$ closest to $T_n^*(x^n)$, when $x^n \in A_n$ and $T_n(x^n) = T_n^*(x^n)$, otherwise.

THEOREM 5.2. *Under* H1, H5 *and* H2 *and* H3 *with* $p = 1$, *we have*:

(i) $\lim_{n\to\infty} T_n(x^{(n)}) = \alpha_0$ *a.s.*
(ii) $(T_n)_{n>k}$ *is ASR at* μ.

PROOF. (i) follows immediately from (5.10), (5.11) and the definition of T_n.
(ii) Let $N = \bigcup_{i=1}^{\infty} N_{1/i}$, where N_ε is previously defined, $N^{*c} = \{x \in R^\infty / T_n^*(x^{(n)}) \to \alpha_0$ and $(T_n^*)_{n>k}$ is resistant at $x\}$ and $M = N \cup N^*$. Given $\varepsilon > 0$ and $x \notin M$ there exist $\delta_1 > 0$ and n_1 such that for all $n \geq n_1$ and

$d_n(x^{(n)}, y^n) \le \delta_1$ the equation $g_n(y^n, \alpha) = 0$ has a root in $[\alpha_0 - \varepsilon/2, \alpha_0 + \varepsilon/2]$. On the other hand, since $x \notin N^*$, there exist $n_2 \ge n_1$ and $\delta_2 < \delta_1$ such that $d_n(x^{(n)}, y^n) \le \delta_2$ and $n \ge n_2$ imply $|T_n^*(y^n) - \alpha_0| < \varepsilon/2$. Therefore, according to the definition of T_n we get (ii). □

Another example of ASR estimates is given by the extension to ARMA models of the S-estimates introduced by Rousseeuw and Yohai (1984) and developed in Boente, Fraiman and Yohai (1985). We will describe them briefly. Let r: $R^{k+1} \times \Theta \to R$ be a continuous function and χ: $R \to R$ be a function verifying E4. Define $\sigma(\theta)$ as the unique solution of

$$f(\theta, \sigma(\theta)) = E\big(\chi\big(r(x_t^k, \theta)/\sigma(\theta)\big)\big) = 0.$$

Let

$$f_n(x^n, \theta, \sigma) = n^{-1} \sum_{t=k+1}^{n} \chi\big(r(x_t^k, \theta)/\sigma\big) = g_n(x^n, \sigma).$$

Then $\sigma_n(\theta, x^n)$ is defined as the unique solution of $g_n(x^n, \sigma_n(\theta, x^n)) = 0$.

We define the S-estimate as the value $\theta_n(x^n)$ such that

$$\sigma_n(x^n) = \sigma_n(\theta_n(x^n), x^n) = \min_\theta \sigma_n(\theta, x^n).$$

Under some regularity conditions, described in Boente, Fraiman and Yohai (1985), it may be shown that the S-estimator is strongly consistent and ASR at μ.

In particular, if we choose

$$r(x_t^k, \theta) = x_t - \hat{x}_t^{(k)}(x_{t-1}^{k-1}, \theta),$$

where $\hat{x}_t^{(k)}(x_{t-1}^{k-1}, \theta)$ is the best linear predictor of x_t based on $x_{t-1}, \dots, x_{t-k}$ when the parameter is θ, and x_t is a Gaussian MA(1) process, the value θ_0 which minimizes $\sigma(\theta)$ is the true parameter. Therefore, this proposal provides consistent and ASR estimates for θ_0.

A class of estimates for AR(p) models which does not depend on any finite dimensional marginal empirical distribution are the RA-estimates considered by Bustos and Yohai (1986), Bustos, Fraiman and Yohai (1984) and Martin and Yohai (1985). It may be shown that under mild conditions these estimates are ASR. The proof is quite similar to the one given above for the GM-estimates.

APPENDIX

The following lemma follows easily from H2 and the ergodic theorem.

LEMMA A.1. *Assume* H1 *and* H2; *then given* $\varepsilon > 0$, *there exists* $\delta > 0$ *such that*

$$\limsup_{n\to\infty} \sup_{|a|=1} \sum_{t=1}^{n} I(|a'z_t| \le \delta)/n \le \varepsilon \quad a.s.$$

LEMMA A.2. *Under* H1, H2 *and* H3, *given* $\varepsilon > 0$, *there exists* $\delta > 0$ *such that*

$$\limsup_{n\to\infty} \sup_{a\in R^p} \sum_{t=1}^{n} I(|y_t - a'z_t| \le \delta)/n \le \varepsilon \quad a.s.$$

PROOF. By Lemma A.1 we can find δ_1 such that

$$\limsup_{n\to\infty} \sup_{|a|=1} \sum_{t=1}^{n} I(|a'z_t| \le 2\delta_1)/n \le \varepsilon/2. \tag{A.1}$$

There exists $k > 1$ such that

$$\mu(|y_t| \ge k\delta_1) \le \varepsilon/2. \tag{A.2}$$

From H3 and the ergodic theorem we obtain

$$\limsup_{n\to\infty} \sup_{|a|\le k} \sum_{t=1}^{n} I(|y_t - a'z_t| \le \delta_2)/n \le \varepsilon \quad \text{a.s.} \tag{A.3}$$

Take $\delta = \min(\delta_1, \delta_2)$. Then by (A.3) it will be enough to prove

$$\limsup_{n\to\infty} \sup_{|a|\ge k} \sum_{t=1}^{n} I(|y_t - a'z_t| \le \delta)/n \le \varepsilon. \tag{A.4}$$

Since $k > 1$, we have

$$\begin{aligned}\sum_{t=1}^{n} I(|y_t - a'z_t| \le \delta)/n &\le \sum_{t=1}^{n} I(|y_t - a'z_t| \le k\delta)/n \\ &\le \sum_{t=1}^{n} I(|y_t| > k\delta)/n + \sum_{t=1}^{n} I(|a'z_t|/k \le 2\delta)/n.\end{aligned}$$

Therefore, by (A.2), the ergodic theorem and (A.1) we have

$$\begin{aligned}&\limsup_{n\to\infty} \sup_{|a|\ge k} \sum_{t=1}^{n} I(|y_t - a'z_t| \le \delta)/n \\ &\quad\le \limsup_{n\to\infty} \sup_{|a|=1} \sum_{t=1}^{n} I(|a'z_t| \le 2\delta)/n + \limsup_{n\to\infty} \sum_{t=1}^{n} I(|y_t| > \delta k)/n \le \varepsilon. \qquad \square\end{aligned}$$

LEMMA A.3. *Assume* E2 *and* E4. *Then given* $d_1 > 0$, $d_2 > 0$ *and* $\varepsilon > 0$ *there exists* $\delta > 0$ *such that for all* $z_1, z_2 \in R^p$, $y_1, y_2 \in R$ *and* $\theta \in R^p$ *satisfying*

$$|z_1 - z_2| \le \delta, \qquad |y_1 - y_2| \le \delta, \qquad d_1 < |z_1| < d_2,$$
$$|y_1| < d_2, \qquad |\theta' z_1|/|\theta| > d_1, \qquad |y_1 - \theta' z_1| > d_1$$

we have

(i) $|\phi(Az_1, r(z_1, y_1, \theta, \sigma))Az_1 - \phi(Az_2, r(z_2, y_2, \theta, \sigma))Az_2| < \varepsilon,$

(ii) $|\chi(r(z_1, y_1, \theta, \sigma)) - \chi(r(z_2, y_2, \theta, \sigma))| < \varepsilon,$

for all $\sigma > 0$.

PROOF. Denote by $r_i = r(z_i, y_i, \theta, \sigma)$, $i = 1, 2$. Then straightforward calculations show that under the stated conditions given $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$|r_1 - r_2|/|\lambda r_1 + (1 - \lambda) r_2| < \varepsilon, \quad \text{for all } \sigma > 0,\ \lambda \in [0,1].$$

Then the desired result follows from E2, E4 and the mean value theorem. □

PROOF OF PROPOSITION 5.1. Let $N^c \subset X^\infty$ be the set where Lemmas A.1 and A.2 hold for all $\varepsilon > 0$ and $C_j = \{(z, y) \in R^p \times R\colon j^{-1} < |z| < j,\ |y| < j\}$. For any $j \geq 1$, the ergodic theorem implies

$$\lim_{n\to\infty} \sum_{t=1}^{n} I(\{x_t \in C_j\})/n = \mu(C_j) \quad \text{a.s.} \tag{A.5}$$

$M^c \subset X$ will denote the set where (A.5) is satisfied for all $j \geq 1$, and $N_1 = N \cup M$. Then $\mu(N_1) = 0$ and in order to prove the proposition it suffices to show that N_1 is the required set.

For any $\varepsilon > 0$, there exists j_0 such that $\mu(C_{j_0}) > 1 - \varepsilon/6D$, where $D = \sup|\varphi(x, \alpha)|$. Given $x \notin N_1$ there exist $\delta_1 \leq 1/\sqrt{p}\, j_0$ and n_1 such that

$$\sup_{|a|=1} \sum_{t=1}^{n} I(|a'z_t| \leq \sqrt{p}\,\delta_1)/n \leq \varepsilon/6D \qquad \forall\, n \geq n_1, \tag{A.6}$$

$$\sup_{a\in R^p} \sum_{t=1}^{n} I(|y_t - a'z_t| \leq \delta_1)/n \leq \varepsilon/6D \qquad \forall\, n \geq n_1. \tag{A.7}$$

Since $x \notin M$, there exists $n_2 \geq n_1$ such that

$$\sum_{t=1}^{n} I(\{x_t \in C_{j_0}\})/n \geq 1 - \varepsilon/6D \qquad \forall\, n \geq n_2. \tag{A.8}$$

For $a \in R^p$ and δ_1 define

$$E_a = \{t\colon 1 \leq t \leq n,\ |a'z_t|/|a| \geq \delta_1,\ |y_t - a'z_t| \geq \delta_1,\ x_t \in C_{j_0}\}^c.$$

From (A.6), (A.7) and (A.8) we obtain $\#E_a/n \leq 4\varepsilon/6D$ for all $n \geq n_2$.

Consider $\delta_0 \leq \varepsilon/6$ as in Lemma A.3 with $d_1 = \delta_1$, $d_2 = j_0$ and $\varepsilon = \varepsilon/6$. Take $w^n \in X^n$ such that $d_n(x^{(n)}, w^n) \leq \delta_0$. Define $F_n = \{t\colon 1 \leq t \leq n\colon |w_t - x_t| \leq \delta_0\}$; therefore, $\#F_n/n \geq 1 - \delta_0 > 1 - \varepsilon/6$. Denote $J = F_n \cap E_\theta^c$; then we have

$$\begin{aligned}|g_n(x^{(n)}, \alpha) - g_n(w^n, \alpha)| \leq \Bigg(&\sum_{t\in J} |\varphi(x_t, \alpha) - \varphi(w_t, \alpha)| \\ &+ \sum_{t\in F_n^c} |\varphi(x_t, \alpha) - \varphi(w_t, \alpha)| \\ &+ \sum_{t\in E_\theta} |\varphi(x_t, \alpha) - \varphi(w_t, \alpha)|\Bigg)\Big/ n.\end{aligned}$$

According to the definition of δ_0, for $t \in J$ we have $|\varphi(x_t, \alpha) - \varphi(w_t, \alpha)| \leq \varepsilon/6$. Then we get that

$$|g_n(x^{(n)}, \alpha) - g_n(w^n, \alpha)| \leq \varepsilon \qquad \forall\, n \geq n_2. \qquad \square$$

Acknowledgments. The authors are very grateful to D. Martin and O. Bustos for their valuable comments and suggestions which have improved the presentation. Thanks are also due D. Cox for helpful discussions and to an anonymous referee who provided a simpler proof of Lemma 2.1.

REFERENCES

BIRKHOFF, G. (1946). Tres observaciones sobre el algebra lineal. *Rev. Univ. Nac. de Tucumán* **5** 147–151.

BOENTE, G., FRAIMAN, R. and YOHAI, V. J. (1982). Qualitative robustness for general stochastic processes. Technical Report 26, Dept. Statistics, Univ. Washington.

BOENTE, G., FRAIMAN, R. and YOHAI, V. J. (1985). Minimum scale estimates for time series models. Unpublished manuscript.

BUSTOS, O. (1981). Qualitative robustness for general processes. Informes de Matemática, Serie B-002/81. Institute de Matemática Pura e Aplicada, Río de Janeiro.

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

BUSTOS, O., FRAIMAN, R. and YOHAI, V. J. (1984). Asymptotic behaviour of the estimates based on residual autocovariances for ARMA models. *Robust and Nonlinear Time Series Analysis. Lecture Notes in Statist.* **26** 26–49. Springer, New York.

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

COX, D. (1981). Metrics on stochastic processes and qualitative robustness. Technical Report 3, Dept. Statistics, Univ. Washington.

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

DONOHO, D. L. (1982). Breakdown properties of multivariate location estimators. Ph.D. qualifying paper, Harvard Univ.

DONOHO, D. L. and HUBER, P. J. (1983). The notion of breakdown point. In *A Festschrift for Erich L. Lehmann* (P. J. Bickel, K. A. Doksum and J. L. Hodges, Jr., eds.) 157–184. Wadsworth, Belmont, Calif.

HAMPEL, F. R. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.* **42** 1887–1896.

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

HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. Fifth Berkeley Symp. Math. Statist. Probab.* **1** 221–233. Univ. California Press.

HUBER, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1** 799–821.

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

KRASKER, W. S. (1980). Estimation in linear regression models with disparate data points. *Econometrica* **48** 1333–1346.

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

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

MALLOWS, C. L. (1976). On some topics in robustness. Bell Labs Technical Memo, Murray Hill, N.J.

MARONNA, R. A. and YOHAI, V. J. (1981). Asymptotic behavior of general M-estimates for regression and scale with random carriers. *Z. Wahrsch. verw. Gebiete* **58** 7–20.

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

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

MOSTELLER, F. and TUKEY, J. W. (1977). *Data Analysis and Regression: A Second Course in Statistics*. Addison-Wesley, Reading, Mass.

PAPANTONI-KAZAKOS, P. and GRAY, R. M. (1979). Robustness of estimators on stationary observations. *Ann. Probab.* **7** 989–1002.

RONCHETTI, E. and ROUSSEEUW, P. J. (1982). Change-of-variance sensitivities in regression analysis. Research Report 36, Fachgruppe für Statistik, ETH Zürich.

ROUSSEEUW, P. and YOHAI, V. (1984). Robust regression by means of S-estimators. *Robust and Nonlinear Time Series Analysis. Lecture Notes in Statist.* **26** 256–272. Springer, New York.

SAMAROV, A. (1983). Bounded-influence regression via local minimax mean square error. Technical Report 40, CCREMS, MIT.

STRASSEN, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statist.* **36** 423–439.

TUKEY, J. N. (1976). Usable resistant/robust techniques of analysis. In *Proc. First ERDA Statistical Symposium, Los Alamos, New Mexico, Nov. 3-5, 1975.* (W. L. Nicholson and J. L. Harris, eds.). Battelle Pacific Northwest Laboratory.

FACULTAD DE CIENCIAS EXACTAS Y NATURALES
DEPARTAMENTO DE MATEMATICA
PABELLON 1, CIUDAD UNIVERSITARIA
1428 BUENOS AIRES
ARGENTINA