Probab. Theory Relat Fields 90, 447–467 (1991)

Probability Theory and Related Fields

The change-of-variance function for dependent data ★

Ola Hössjer

Department of Mathematics, Uppsala University, S-75238 Uppsala, Sweden

Received March 28, 1990; in revised form June 20, 1991

Summary. The infinitesimal stability of the asymptotic variance is considered for M-estimators of a location parameter when the nominal sample with i.i.d. data is contaminated by a possibly dependent process. It is shown that the resulting change-of-variance function can be expressed as a sum of two terms, one corresponding contamination of the univariate distribution, and one to contamination of the bivariate distributions. A change-of-variance sensitivity is introduced, the form of which is closely related to the average patch length of the outliers. Finally, optimal V-robust and most V-robust score functions are derived. The resulting family of estimators is the same as for independent data in the general case, but the truncation point approaches zero when dependency is accounted for. For redescending score-functions, the family of estimators is changed.

1 Introduction

M-estimators of location for independent and identically distributed (i.i.d.) data were introduced by Huber (1964), where he studied their minimax robustness properties. The infinitesimal robustness of the asymptotic value was investigated by Hampel (1974) by means of the influence curve and later on Rousseeuw (1981) introduced the change-of-variance curve, describing the robustness of the asymptotic variance.

The independence assumption in statistical data is often violated. Robustness against serial correlation is therefore a desirable property of a statistical procedure. However, generalizations of infinitesimal robustness concepts to time series impose difficulties since the functionals describing the estimates do no longer depend on the one-dimensional marginal distribution of the data only. Künsch (1984) generalized Hampel's influence curve to M-estimates of AR-parameters, where the functional depends on a finite-dimensional marginal distribution of the underlying process. An even more gen-

★ This paper was written under support by the Swedish Board for Technical Development, contract 712-89-1073

eral notion of influence curve for estimates depending on the marginal distributions of all orders (such as M-estimates of MA-parameters) was introduced by Martin and Yohai (1986a). In that paper they also described the outlier configuration in a rather general way through a replacement model. This model makes it possible to distinguish between outliers occuring individually and in patches.

The purpose of this paper is to use Martin and Yohai's uncertainty model (in a somewhat more general form) in order to generalize Rousseeuw's change-of-variance curve CVF (Sect. 3) and change-of-variance κ^* (Sect. 5) to dependent data. We restrict ourselves to M-estimators of a single location parameter and the data are assumed to be nominally independent. Both CVF and κ^* contain additional terms compared to the independent case, and these terms can be related to the distribution of the patch length of the outliers in a rather simple way. A quantity α, which we call the infinitesimal average patch length, is of fundamental importance here. In Sect. 6 we derive optimal V-robust M-estimators by minimizing the asymptotic variance subject to an upper bound constraint k on κ^*. In the general case (Subsect. 6.1), the resulting family of M-estimators is the same as for independent data, whereas for redescending M-estimators (Subsect. 6.2) it is different. However, given a fixed value of k, the resulting optimal V-robust estimator depends on the amount of dependency allowed for, even in the general case. Finally, in Sect. 7 we outline some possible generalizations of the studied model.

The special case of nominal independence treated here is important, since serial correlations is often an unsuspected (or at least unwanted) feature of the data, and should be regarded as a deviation from the assumed parametric model.

Moustakides and Thomas (1984), Sadowsky (1986) and Zamar (1990) have investigated the stability of the asymptotic variance against dependence for M-estimators of location. They use minimax techniques, and their uncertainty classes are based on ε-contamination for the univariate marginal distribution and various mixing conditions for the bivariate distributions. Portnoy (1977, 1979) considers M-estimators of location with moving average type errors having a weak correlation structure, and he obtains (approximate) minimax solutions for the asymptotic variance. Lee and Martin (1986) treat the ARMA model and compare (ordinary) M-estimators of location with proper M-estimators (i.e., joint estimation of the location parameter and the ARMA-parameters), and show that the ordinary estimator can be quite inefficient compared to the proper one, when the correlation structure is moderate to large. As mentioned above, only infinitesimal robustness of the bias has been studied so far in the time series setting. However, there are situations when the asymptotic variance is a more relevant performance criterion. For instance, the Pitman efficacy is a useful quantity in testing situations (cf. Noether 1955) which is closely related to (for M-estimators just the inverse of) the asymptotic variance. Let us also mention that with our contamination model, the variance of the estimator always tends to zero as $1/n$ (at least formally) with increasing sample size n. In some cases, $1/n^\alpha$, $0<\alpha<1$ is a more realistic convergence rate (i.e., when so called semi-systematic errors occur). One way of modelling such data is by means of self-similar processes (cf. Mandelbrot 1977). Parameter estimation in such long-range dependence processes is treated by e.g. Beran and Künsch (1985), Fox and Taqqu (1986) and Künsch (1987).

2 Preliminaries

2.1 M-estimators.

Let $y_1, \ldots, y_n$ be observations of a stationary ergodic process $\{Y_i\}_{i=-\infty}^{\infty}$ with associated measure μ_y on $(R^{-\infty,\infty}, \mathscr{B}^{-\infty,\infty})$ and let F_y be the univariate distribution of the process. The location parameter θ is defined as the solution of the equation

$$(2.1) \qquad \int \psi(\xi-\theta)\,\mathrm{d}F_y(\xi)=0,$$

given some function ψ. We define the functional $T\colon \mu_y \to T(\mu_y)$ by putting $T(\mu_y) = \theta$. An M-estimate T_n of the location parameter θ is a solution of the equation

$$(2.2) \qquad \sum_{i=1}^{n} \psi(y_i - T_n)=0.$$

Suppose from now on that $T(\mu_y)=0$. (This will be the case for all processes whose asymptotic variance we consider.) Under suitable regularity conditions we then have $T_n \xrightarrow{p} 0$ (consistency) and that $\sqrt{n}\,T_n$ is asymptotically normal with zero mean and variance

$$(2.3) \qquad V(\psi,\mu_y)=\frac{\int \psi(\xi)^2\,\mathrm{d}F_y(\xi)+2\sum_{k=1}^{\infty}\iint \psi(\xi_1)\,\psi(\xi_2)\,\mathrm{d}F_y^{(k)}(\xi_1,\xi_2)}{\left(\int \psi'(\xi)\,\mathrm{d}F_y(\xi)\right)^2}$$

where $F_y^{(k)}$ is the bivariate distribution of the pair (Y_1, Y_{1+k}). Regularity conditions for consistency and asymptotic normality are given by Huber (1967) for i.i.d. data and by Portnoy (1977) and Bustos (1982) for dependent data. In order to prove asymptotic normality, some mixing condition is needed, e.g. that $\{Y_i\}$ is ϕ-mixing with $\sum_{i=1}^{\infty} \phi_i^{1/2} < \infty$, cf. Billingsley (1968, Sect. 21).

2.2 Uncertainty model.

The contaminated process is constructed according to the following general replacement model:

$$(2.4) \qquad Y_i^\gamma=(1-Z_i^\gamma)\,X_i+Z_i^\gamma\,W_i,$$

where $0 \leqq \gamma \leqq 1$, X_i is the nominal i.i.d. process, W_i the contaminating process and Z_i^γ a 0-1 process with

$$(2.5) \qquad P(Z_i^\gamma=1)=g_0(\gamma)=\gamma+o(\gamma).$$

This uncertainty model has been used by Martin and Yohai (1986a). We denote by μ_x, μ_w, μ_z^γ and μ_y^γ the measures on $(R^{-\infty,\infty}, \mathscr{B}^{-\infty,\infty})$ corresponding to the processes above.

According to (2.4), the measure μ_y^γ is determined by the joint measure μ_{xwz}^γ. When Z_i^γ and W_i are i.i.d. processes, we obtain the ordinary contamination model for independent data. By introducing dependency in the Z_i^γ- and W_i-processes, it is possible to model patchy outliers. Zamar (1990, Lemma 1) also considers a contamination model of the kind (2.4), with Z_i^γ an i.i.d. process.

2.3 Notation and regularity conditions.

First some notation: We let F_x and F_y^γ denote the univariate distribution of the X_i- and Y_i^γ-processes and F_w^γ the distribution of W_i conditioned on the event $\{Z_i^\gamma=1\}$. The bivariate distribution of the pair $(Y_i^\gamma, Y_{i+k}^\gamma)$ will be written $F_y^{\gamma,(k)}$ and that of (W_i, W_{i+k}), conditioned on the event $\{Z_i^\gamma=Z_{i+k}^\gamma=1\}$, we denote by $F_w^{\gamma,(k)}$. We also write $F_{xw}^{\gamma,(k)}$ and $F_{wx}^{\gamma,(k)}$ for the bivariate distributions of (X_i, W_{i+k}) and (W_i, X_{i+k}), conditioned on the events $\{Z_i^\gamma=0, Z_{i+k}^\gamma=1\}$ and $\{Z_i^\gamma=1, Z_{i+k}^\gamma=0\}$ respectively. In cases when the conditional W-distributions are the same for all values of γ (which will be assumed throughout the paper except for Subsect. 7.1), we will drop γ as superscript and simply write F_w and $F_w^{(k)}$.

The nominal univariate distribution F_x is kept fixed and satisfies:

(A1) F_x has a twice continuously differentiable density f_x which is symmetric and strictly positive.

(A2) The maximum likelihood function $\Lambda=-f_x'/f_x$ satisfies $\Lambda'(\xi)>0$ for all ξ, and $\int \Lambda'(\xi)\, f_x(\xi)\,\mathrm{d}\xi=-\int \Lambda(\xi)\, f_x'(\xi)\,\mathrm{d}\xi<\infty$.

(A1)–(A2) imply that F_x has finite Fisher information and that f_x is unimodal. Moreover, all regularity conditions on the nominal distribution in the minimax asymptotic variance theorem in (Huber 1964, p. 80) are satisfied.

The class Ψ of all admissible functions ψ is specified through

(B1) ψ is continuous on $R\backslash C(\psi)$, where $C(\psi)$ is finite. In each point of $C(\psi)$, there exist finite left and right limits of ψ which are different. Furthermore, $\psi(-\xi)=-\psi(\xi)$ if $\{\xi, -\xi\}\subset R\backslash C(\psi)$ and $\psi(\xi)\geqq 0$ if $\xi\geqq 0$ and $\xi\in R\backslash C(\psi)$.

(B2) The set of points $D(\psi)$ where ψ is continuous but ψ' is not defined or not continuous, is finite.

(B3) $\int \psi(\xi)^2\,\mathrm{d}F_x(\xi)<\infty$.

(B4) $0<\int \psi'(\xi)\,\mathrm{d}F_x(\xi)=-\int \psi(\xi)\, f_x'(\xi)\,\mathrm{d}\xi=\int \Lambda(\xi)\,\psi(\xi)\,\mathrm{d}F_x(\xi)<\infty$.

By allowing ψ to have a finite number of discontinuities, a large class of M-estimators will be contained in the class Ψ, including the median, the Huber-type skipped median and the median-type tanh-estimator. Since ψ may have discontinuity points, ψ' is to be interpreted as a Schwarz distribution, as described by Rousseeuw (1982), so that

$$\int \psi'(\xi)\,\mathrm{d}F(\xi)=\int_{R\backslash(C(\psi)\cup D(\psi))} \psi'(\xi)\,\mathrm{d}F(\xi)+\sum_{i=1}^{m} (\psi(c_i+)-\psi(c_i-))\, f(c_i), \tag{2.6}$$

where the first term is the classical integral, $c_1<\ldots<c_m$ are the points of $C(\psi)$ and f is the density (assumed to be continuous in a neighbourhood of $C(\psi)$) of an arbitrary distribution F. Conditions (A1)–(A2) and (B1)–(B4) are the same as those given by Rousseeuw (1982).

Next, we make the following assumptions about the outlier generating processes Z_i^γ and the contamination process W_i:

(C1) The Z_i^γ-processes are independent of the nominal process and $P(Z_1^\gamma=Z_{1+k}^\gamma=1)=g_k(\gamma)=\alpha_k\,\gamma+r_k(\gamma)$, $k=0, 1, 2, \ldots$, with $r_k(\gamma)=o(\gamma)$ and $\sum_{k=0}^{\infty} \alpha_k<\infty$.

(C2) X_i and W_j are conditionally independent on the event $\{Z_i^\gamma=0, Z_j^\gamma=1\}$ when $i \neq j$. (However, when $i=j$ the two random variables may be dependent and consequently, the processes may be dependent.)

(C3) The conditional distributions F_w^γ and $F_w^{\gamma,(k)}$ are independent of γ (and hence will be written F_w and $F_w^{(k)}$ respectively).

(C4) F_w has a symmetric density f_w, $\int \psi(\xi)^2 \, \mathrm{d}F_w(\xi) < \infty$ and $\sum_{k=1}^{\infty} |\iint \psi(\xi_1)\,\psi(\xi_2)\,\mathrm{d}F_w^{(k)}(\xi_1, \xi_2)| < \infty$.

(C5) The integral $\int \psi'(\xi)\,\mathrm{d}F_w(\xi)$ exists in the sense of (2.6) and is finite.

Condition (C1) expresses the bivariate dependency structure of the outlier generating process Z_i^γ. Comparing (C1) with (2.5), we observe that $\alpha_0=1$. We will see in Sect. 4, that the constants α_k are closely related to the distribution of the outlier patch length. In particular, for i.i.d. data (with isolated outliers) we have $\alpha_k=0$ when $k \neq 0$, since $\alpha_k = \lim_{\gamma \to 0} P(Z_{1+k}^\gamma=1 \mid Z_1^\gamma=1)$. Condition (C2) guarantees that all mixed terms of the kind $E(\psi(X_i)\,\psi(W_j) \mid Z_i^\gamma=0, Z_j^\gamma=1)$ vanish in the expression for the asymptotic variance. Intuitively, this condition can be interpreted by the statement that the distribution of an outlier occuring at time j which is not present at time i is independent of what happens at time i. (C2) is valid for additive outlier models such as in Subsect. 7.1. Condition (C3) obviously includes the case when the Z_i^γ- and W_i-processes are independent. In general, when the two processes are dependent, it is clear that the conditional W-distributions may vary with γ. In that case one could assume that F_w^γ and $F_w^{\gamma,(k)}$ converge weakly to some limiting distributions F_w and $F_w^{(k)}$ as $\gamma \to 0$. Such a convergence is established for the additive outlier model in Subsect. 7.1. It turns out then that formula (3.2) for the change-of-variance function still holds true. However, this requires more regularity on ψ (see Theorem 7.1), leaving out many important estimators. We therefore prefer to work with (C3) in the following. It is also important to observe that in general the one-dimensional marginal distributions of $F_w^{(k)}$ need not equal F_w, since the two distributions are conditioned on different events (see Subsect. 7.1). Conditions (C4)–(C5) guarantee the existence of the change-of-variance function in (3.1). Furthermore, the symmetry of f_x and f_w together with the antisymmetry of ψ imply that $\theta=0$ is a solution of (2.1) for the contaminated process Y_i^γ.
Finally, we introduce the notation

$$A(\psi) = \int \psi(\xi)^2 \, \mathrm{d}F_x(\xi) \tag{2.7}$$

and

$$B(\psi) = \int \psi'(\xi) \, \mathrm{d}F_x(\xi) \tag{2.8}$$

so that the nominal asymptotic variance is given by

$$V(\psi, \mu_x) = A(\psi)/B(\psi)^2. \tag{2.9}$$

2.4 Infinitesimal bias. According to (2.2), the infinitesimal variations of the asymptotic value of the M-estimator only depends on the univariate distribution of the contaminating process. By differentiating (2.1) with respect to γ one obtains (Hampel 1974; Martin and Yohai 1986a)

$$\frac{\partial}{\partial\gamma}[T(\mu_y^\gamma)]_{\gamma=0}=\int IF(\psi, F_x, \xi)\,\mathrm{d}F_w(\xi), \tag{2.10}$$

where IF is the influence function, which is given by

$$IF(\psi, F_x, \xi)=\psi(\xi)/B(\psi), \tag{2.11}$$

on $R\backslash C(\psi)$. The maximal infinitesimal bias is described by the gross-error sensitivity

$$\gamma^*(\psi, F_x)=\sup_{\xi\in R\backslash C(\psi)}|IF(\psi, F_x, \xi)|. \tag{2.12}$$

It is sometimes more convenient to work with the standardized gross-error sensitivity

$$\gamma_s^*(\psi, F_x)=\gamma^*(\psi, F_x)/\sqrt{V(\psi, F_x)} \tag{2.13}$$

which is invariant with respect to affine transformations of the parameter space.

3 Change-of-variance function

Having defined the uncertainty model, we are now ready to introduce the change-of-variance function (CVF), which we define in the following way:

$$\mathrm{CVF}(\psi, \{\mu_y^\gamma\})=\frac{\partial}{\partial\gamma}[\log V(\psi, \mu_y^\gamma)]_{\gamma=0}. \tag{3.1}$$

Note that according to this definition, the CVF does not only depend on ψ, μ_x and the contaminating measure μ_w, but also on the trajectory of contaminated measures $\{\mu_y^\gamma, 0\leqq\gamma\leqq1\}$. By differentiating formula (2.3) with respect to γ we obtain the following expression for the CVF:

Theorem 3.1 *Assume that the regularity conditions of Subsect.* 2.3 *hold. The* CVF *defined in* (3.1) *is then given by*

$$\mathrm{CVF}(\psi, \{\mu_y^\gamma\})=1+\frac{\int\psi(\xi)^2\,\mathrm{d}F_w(\xi)+2\sum_{k=1}^{\infty}\alpha_k\iint\psi(\xi_1)\,\psi(\xi_2)\,\mathrm{d}F_w^{(k)}(\xi_1, \xi_2)}{A(\psi)} -2\frac{\int\psi'(\xi)\,\mathrm{d}F_w(\xi)}{B(\psi)}. \tag{3.2}$$

For i.i.d. data, with $\alpha_k=0$ when $k\neq 0$, (3.2) reduces to the expression

$$\text{(3.3)} \qquad \mathrm{CVF}(\psi, \{\mu_y^\gamma\})=1+\frac{\int \psi(\xi)^2 \,\mathrm{d}F_w(\xi)}{A(\psi)}-2\frac{\int \psi'(\xi)\,\mathrm{d}F_w(\xi)}{B(\psi)},$$

derived by Rousseeuw (1981).

4 Interpretation of the outlier generating process

In this section, we will motivate and interpret the regularity conditions imposed on the Z_i^γ-process in (C1) of the previous section. Given that an outlier occurs at a certain time, let T be a positive, integer-valued random variable describing the duration of the outliers, with

$$\text{(4.1)} \qquad P(T>j)=\theta_j, \quad j=0,1,2,\ldots$$

Assume that the probability of a (first) occurence of an outlier at a certain time is $p, 0\leqq p\leqq 1$ and let the random variable $\tilde{Z}_{j,i}^p$, $j\leqq i$ be the indicator for the event that an outlier first occuring at time j is still present at time i. According to (4.1), we then have

$$\text{(4.2)} \qquad P(\tilde{Z}_{j,i}^p=1)=p\theta_{i-j}.$$

Since Z_i^γ (where γ will be related to p below) is the indicator of the event that an outlier occurs at time i we have

$$\text{(4.3)} \qquad Z_i^\gamma=1-\prod_{j\leqq i}(1-\tilde{Z}_{j,i}^p).$$

We will also assume that

(D1) $E(T^2)<\infty$.

(D2) The random variables $\tilde{Z}_{j,i}^p$, $j\leqq i$ are jointly independent of the X_i-process and all sequences $\{\tilde{Z}_{j,i_j}^p\}_{j=-\infty}^{\infty}$ of random variables are independent.

Condition (D2) simply says that the time durations of outliers occuring (first) at different times are independent and furthermore, they are independent of the nominal process. It is now possible to show that the regularity conditions imposed on the Z_i^γ-process in Subsect. 2.3 are satisfied for the particular Z_i^γ-process generated here:

Theorem 4.1 *Suppose that the family of outlier generating processes $Z_i^\gamma, 0\leqq\gamma\leqq 1$, satisfies* (4.3) *for each γ and that* (D1)–(D2) *hold. Then condition* (C1) *of Sect.* 2 *holds with parameters γ and α_k, $k=0,1,2,\ldots$ given by*

$$\text{(4.4)} \qquad \gamma=p\sum_{j=0}^{\infty}\theta_j=pE(T)$$

and

$$\text{(4.5)} \qquad \alpha_k=\sum_{j=k}^{\infty}\theta_j \Big/ \sum_{j=0}^{\infty}\theta_j.$$

Remark. In order to show that conditions (C2)–(C5) are satisfied, we have to choose a specific contamination process W_i.

As an alternative description of the configuration of the outliers, we introduce $\tilde{L}_i^\gamma$ as the total length of the (possible) patch of outliers at time i. If $Z_i^\gamma=0$, we let $\tilde{L}_i^\gamma=0$, and if $Z_i^\gamma=1$, put

$$\tilde{L}_i^\gamma=\max\{|k-j+1|; j\leq i\leq k, Z_j^\gamma=Z_{j+1}^\gamma=\ldots=Z_k^\gamma=1\}. \tag{4.6}$$

Furthermore, let

$$L_i^\gamma=\tilde{L}_i^\gamma|Z_i^\gamma=1 \tag{4.7}$$

describe the patch length of outliers conditioned on the event that an outlier is present. We then have the following result:

Theorem 4.2 *Given the same assumptions as in Theorem* 4.1, *there exists a positive, integer valued random variable L such that* $L_i^\gamma \xrightarrow{d} L$ *for all i,*

$$P(L=j)=\frac{jP(T=j)}{E(T)}, \quad j=1,2,\ldots, \tag{4.8}$$

$L\overset{s}{\geq}T$ *(i.e., is stocastically larger than), with equality iff* $T=\delta_l$ *for some l, and finally,*

$$E(L)=\frac{ET^2}{ET}=\alpha=1+2\sum_{k=1}^{\infty}\alpha_k. \tag{4.9}$$

The quantity α, which we call the "infinitesimal average patch length", turns out to be of great importance as a measure of dependence when calculating the change-of-variance sensitivity in Sect. 5 and deriving V-robust estimators in Sect. 6.

Example 4.1 As an illustration, assume that all outliers have patch length l, i.e. $T=\delta_l$. In this case the Z_i^γ-process is given by

$$Z_i^\gamma=1-\prod_{j=i-l+1}^{i}(1-\tilde{Z}_{j,i}^p), \tag{4.10}$$

with $\gamma=lp$, $\alpha_k=1-k/l$ when $0\leq k<l$ and $\alpha_k=0$ when $k\geq l$. Moreover, $L=\delta_l$ and $\alpha=l$. This uncertainty model was used by Martin and Yohai (1986a) for constructing patchy outliers.

Example 4.2 If T has a geometric distribution with parameter r, then $\gamma=p/r$ and $\alpha_k=(1-r)^k$. Furthermore, $L+1$ has a negative binomial distribution with parameters 2 and r, and $\alpha=\dfrac{2}{r}-1$.

5 Change-of-variance sensitivity

In this section we give an upper bound for the CVF given by (3.2) when the trajectory of contamination measures is allowed to vary. This is achieved by keeping μ_x and $\{\mu_z^\gamma, 0 \leqq \gamma \leqq 1\}$ fixed and letting μ_w vary. In order to obtain a simple expression for the upper bound of CVF, the supremum will be taken over a class $\mathscr{W}$ of contamination measures μ_w whose members satisfy (C2)–(C5) of Sect. 3, and furthermore, the one-dimensional marginals of $F_w^{(k)}$ equal F_w for each k. The latter constraint is satisfied for all contamination processes that are independent of the Z_i^γ-processes for all γ. Note that in particular, formula (3.2) is valid for each element of $\mathscr{W}$.

For a fixed μ_w, we may apply Cauchy-Schwarz inequality on each term of the infinite series in (3.2) to obtain an upper bound for CVF:

$$\text{(5.1)} \qquad \mathrm{CVF}(\psi, \{\mu_y^\gamma\}) \leqq \int \left(1 + \frac{\alpha \psi(\xi)^2}{A(\psi)} - \frac{2\psi'(\xi)}{B(\psi)}\right) \mathrm{d}F_w(\xi),$$

where α is the infinitesimal average patch length defined in (4.9). Now taking the supremum in (5.1) over $\mathscr{W}$ and remembering the interpretation (2.6) of the integration of ψ', we have

$$\text{(5.2)} \qquad \sup_{\mu_w \in \mathscr{W}} \mathrm{CVF}(\psi, \{\mu_y^\gamma\}) = \kappa_\alpha^*(\psi, F_x)$$

where we define the change-of-variance sensitivity $\kappa_\alpha^*(\psi, F_x)$ as follows:

Definition. The change-of-variance sensitivity $\kappa_\alpha^*(\psi, F_x)$ is defined as $+\infty$ if a delta function with negative factor occurs in ψ', and otherwise as

$$\text{(5.3)} \qquad \kappa_\alpha^*(\psi, F_x) = \sup\left\{1 + \frac{\alpha \psi(\xi)^2}{A(\psi)} - \frac{2\psi'(\xi)}{B(\psi)}, \xi \in R \setminus (C(\psi) \cup D(\psi))\right\}.$$

This definition generalizes that of Rousseeuw (1982) for independent data, which corresponds to the case $\alpha = 1$.

In order to show that $\kappa_\alpha^*(\psi, F_x)$ is a tight upper bound for the supremum in (5.2) we give the following example. Let U_i and V_i be processes jointly independent of the X_i- and Z_i^γ-processes. The U_i- and V_i-processes are also mutually independent i.i.d. processes with univariate distribution $Be(p)$ and F_v respectively. The contamination process is now defined as a Markov process:

$$\text{(5.4)} \qquad W_i = \begin{cases} V_i, & \text{if } U_i = 0 \\ W_{i-1}, & \text{if } U_i = 1. \end{cases}$$

This process has univariate distribution $F_w = F_v$ and bivariate distributions $F_w^{(k)} = (1-p^k) F_v \times F_v + p^k \delta_{F_v}$, where δ_{F_v} is a diagonal measure, defined in such a way that $\iint h(\xi_1, \xi_2)\, \mathrm{d}\delta_{F_v}(\xi_1, \xi_2) = \int h(\xi, \xi)\, \mathrm{d}F_v(\xi)$ for any bounded and continuous function h. If now F_v has a symmetric density satisfying $\int \psi(\xi)^2 \mathrm{d}F_v(\xi) < \infty$ and

$|\int \psi'(\xi)\,\mathrm{d}F_v(\xi)| < \infty$ it follows that conditions (C2)–(C5) of Subsect. 2.3 are satisfied with change-of-variance function

$$\text{(5.5)} \qquad \mathrm{CVF}(\psi, \{\mu_y^\gamma\}) = 1 + \alpha(p)\frac{\int \psi(\xi)^2\,\mathrm{d}F_v(\xi)}{A(\psi)} - 2\frac{\int \psi'(\xi)\,\mathrm{d}F_v(\xi)}{B(\psi)},$$

where

$$\text{(5.6)} \qquad \alpha(p) = 1 + 2\sum_{k=1}^{\infty} p^k \alpha_k.$$

By letting $p \to 1$ and varying F_v we see that the supremum in (5.2) must equal $\kappa_\alpha^*(\psi, F_x)$.

Remark. In view of (5.1) and (5.3), it is natural to introduce a change-of-variance curve (CVC) in the following way:

$$\text{(5.7)} \qquad \mathrm{CVC}(\psi, \xi) = 1 + \frac{\alpha\psi(\xi)^2}{A(\psi)} - \frac{2\psi'(\xi)}{B(\psi)},$$

where the last term may contain delta functions as in (2.6). Let $\bar{\xi} = (\ldots, \xi, \xi, \xi, \ldots)$ be a constant sequence in $R^{-\infty,\infty}$. Then formally, $\mathrm{CVC}(\psi, \xi)$ corresponds to choosing $\mu_w = \frac{1}{2}\delta_{\bar{\xi}} + \frac{1}{2}\delta_{-\bar{\xi}}$ in (3.2), with μ_w independent of μ_z^γ. Unfortunately, this process does not satisfy (C4), so the CVF does not exist. To remedy this, we can let μ_ω^p be the Markov process above with $F_v = \frac{1}{2}\delta_\xi + \frac{1}{2}\delta_{-\xi}$. Letting $\{\mu_y^{p,\gamma}, 0 \leqq \gamma \leqq 1\}$ be the corresponding family of contaminated measures we obtain

$$\text{(5.8)} \qquad \mathrm{CVC}(\psi, \xi) = \lim_{p \to 1} \mathrm{CVF}(\psi, \{\mu_y^{p,\gamma}\}).$$

6 V-Robustness

6.1 General case.

In this section, we will state a number of optimality properties related to $\kappa_\alpha^*(\psi, F_x)$, given a fixed α. Unless otherwise stated, the results are straightforward generalizations of the case $\alpha = 1$ and proved in the same way (see Rousseeuw 1982).

Following the notation of Rousseeuw (1981) we say that an M-estimator with corresponding ψ-function is B-robust if $\gamma^*(\psi, F_x) < \infty$ (which is equivalent to $\gamma_s^*(\psi, F_x) < \infty$, since $0 < V(\psi, F_x) < \infty$). Given a fixed α, we call the M-estimator V_α-robust if $\kappa_\alpha^*(\psi, F_x) < \infty$. Putting $\|\psi\| = \sup_{\xi \in R} \psi(\xi)$ we then have:

Theorem 6.1 *For all ψ in Ψ, V_α-robustness implies B-robustness, and*

$$\text{(6.1)} \qquad 1 + \alpha\gamma_s^*(\psi, F_x)^2 = 1 + \frac{\alpha\|\psi\|^2}{A(\psi)} \leqq \kappa_\alpha^*(\psi, F_x),$$

with equality if ψ is nondecreasing. Moreover, V_α-robustness is equivalent to V_1-robustness, with

$$(6.2)\qquad \kappa_1^*(\psi, F_x) \leqq \kappa_\alpha^*(\psi, F_x) \leqq \alpha(\kappa_1^*(\psi, F_x) - 1) + 1,$$

and the last inequality can be replaced by equality if ψ is nondecreasing.

A frequently used class of robust estimators are the truncated maximum-likelihood estimators, corresponding to

$$(6.3)\qquad \psi_b(\xi) = [\Lambda(\xi)]_{-b}^{b},$$

where $0 < b < \|\Lambda\| = \sup_{\xi \in R} |\Lambda(\xi)|$. The limiting case $b \to 0$ corresponds to the sample median, with score function

$$(6.4)\qquad \psi_{\text{med}}(\xi) = \text{sgn}(\xi).$$

We say that two estimators corresponding to ψ_1 and ψ_2 are equivalent if $C(\psi_1) = C(\psi_2)$ and $\psi_1(\xi) = c\psi_2(\xi)$, for all $\xi \in R \backslash C(\psi_1)$, with c a fixed constant. An estimator will be called optimal V_α-robust if it minimizes $V(\psi, \mu_x)$ among all $\psi \in \Psi$ satisfying the constraint $\kappa_\alpha^*(\psi, \mu_x) \leqq k$ for some fixed number $k < \infty$ and most V_α-robust if it minimizes $\kappa_\alpha^*(\psi, \mu_x)$ in the whole class Ψ. These definitions generalize the notions optimal V-robust and most V-robust estimators (Rousseeuw 1981), for the case $\alpha = 1$. The next theorem states that the optimal V_α-robust and most V_α-robust estimators are the same for all values of α.

Theorem 6.2 *Up to equivalence, the median is the unique most V_α-robust estimator in Ψ with $\kappa_\alpha^*(\psi_{\text{med}}, \mu_x) = \alpha + 1$ and the only optimal V_α-robust estimators are (depending on k) $\psi_{\text{med}}, \{\psi_b, 0 < b < \|\Lambda\|\}$ and Λ if $\|\Lambda\| < \infty$.*

We see from Theorem 6.2 that the family of optimal V_α-robust estimators is the same for all values of α, and hence the same as for independent data ($\alpha = 1$). However, given a fixed constraint $\kappa_\alpha^* \leqq k$ (with $k \geqq \alpha + 1$) the truncation point $b = b(\alpha, k)$ clearly depends on α, and it is easy to see that with k fixed, $b(\alpha, k)$ is a strictly decreasing function of α (since $\kappa_\alpha^*([\psi]_{-b}^{b}, \mu_x)$ is a strictly increasing function of b, cf. Rousseeuw 1981). Hence, for large values of the infinitesimal average patch length α, the optimal V_α-robust score functions approach the sample median. This is illustrated in Table 1, when F_x equals the standard normal distribution and $\psi_b(\xi) = [\xi]_{-b}^{b}$ corresponds to a Huber estimator. A similar conclusion is derived by Zamar (1990, Theorem 1 and Table 2) for minimax robustness.

Table 1. Optimal choices of truncation point b for various values of α and k, the upper bound constraint on κ_α^*

α	$k=2$	$k=3$	$k=5$	$k=10$	$k=20$
1	0.00	1.04	1.90	2.99	4.36
2	–	0.00	1.04	2.04	3.08
3	–	–	0.48	1.55	2.49
4	–	–	0.00	1.19	2.11
5	–	–	–	0.90	1.83
8	–	–	–	0.21	1.26
15	–	–	–	–	0.40

Optimal V_α-robustness is a compromise between low values of $V(\psi, \mu_x)$ (nominal performance) and $\kappa_\alpha^*(\psi, F_x)$. As α becomes large, the second quantity becomes more important in this trade-off and the optimal V_α-robust choice of ψ approaches the most V_α-robust score function. This explains heuristically why a more dependent contamination requires a more V-robust estimator.

6.2 Redescending M-estimators.

As we saw in the preceding subsection, the V-robustness properties for the whole class Ψ of M-estimators extended directly from the case $\alpha = 1$. However, if we restrict ourselves to the subclass of M-estimators with rejection point r,

$$\Psi_r = \{\psi \in \Psi, \psi(\xi) = 0, \text{for all } |\xi| > r\}, \tag{6.5}$$

the situation is slightly different. It turns out that the most V_α-robust and the family of optimal V_α-robust estimators ψ depends on α. The case $\alpha = 1$ is treated by Rousseeuw (1982). In general, the optimal V_α-robust estimators within the subclass (6.5) are estimators with score functions of the form

$$\chi_{r,\alpha,k}(\xi) = \begin{cases} A(\xi), 0 \leqq |\xi| \leqq p \\ \sqrt{\dfrac{A(k-1)}{\alpha}} \tanh\left(\dfrac{1}{2} B \sqrt{\dfrac{\alpha(k-1)}{A}} (r - |\xi|)\right) \operatorname{sgn}(\xi), p < |\xi| \leqq r \\ 0, |\xi| > r, \end{cases} \tag{6.6}$$

whenever k, the upper bound constraint on κ_α^*, exceeds a certain lower bound $\kappa_{r,\alpha}$. In (6.6), $A = A(\chi_{r,\alpha,k})$, $B = B(\chi_{r,\alpha,k})$ and p are chosen so that $\chi_{r,\alpha,k}$ becomes continuous. The family $\{\chi_{r,\alpha,k}\}_{k=\kappa(r,\alpha)}^{\infty}$ corresponds to the tanh-estimators derived by Rousseeuw (1982) when $\alpha = 1$, and to the family of optimal B-robust estimators as $\alpha = \infty$. The latter one consists of skipped Huber estimators $\{\psi_{r,b}\}_{b=0}^{A(r)}$, where

$$\psi_{r,b}(\xi) = [A(\xi)]_{-b}^{b} \, 1_{[-r,r]}(\xi), \tag{6.7}$$

$b = 0$ corresponds to the skipped median and $b = A(r)$ to the Huber-type skipped mean (cf. Rousseeuw 1982). As $\alpha \to \infty$, the last term in (5.3) becomes negligible, and the corresponding minimization problem approaches that for optimal B-robustness. For details, cf. Hössjer (1989).

7 Generalizations

7.1 Generalized uncertainty model.

Consider the following additive outlier model, proposed by Franke and Hannan (1986):

$$Y_i^\gamma = X_i + \sum_{j=0}^{l-1} \beta_j V_{i-j}^p, \tag{7.1}$$

where $\{V_i^p\}_{i=-\infty}^{\infty}$ are i.i.d. random variables jointly independent of the nominal process, with distribution $(1-p)\delta_0 + pH$, $0 \leqq p \leqq 1$ (the value of γ will be specified below), H is a symmetric distribution with point mass zero at the origin and

$\beta_0, \ldots, \beta_{l-1}$ are nonzero constants. This family of processes can be described by a replacement model that is slightly more general than (2.4) (see also Martin and Yohai 1986b):

$$Y_i^\gamma = (1 - Z_i^\gamma)\, X_i + Z_i^\gamma\, W_i^\gamma. \tag{7.2}$$

The meaning of the 0-1 process Z_i^γ becomes clear when comparing (7.1) with (7.2), we simply put

$$\{Z_i^\gamma = 1\} = \bigcup_{j=0}^{l-1} \{V_{i-j}^p \neq 0\}. \tag{7.3}$$

By defining

$$\tilde{Z}_{j,i}^p = \begin{cases} I(V_j^p \neq 0), & i-l < j \leqq i \\ 0, & j \leqq i-l, \end{cases} \tag{7.4}$$

we see that (7.3) can be rewritten as

$$Z_i^\gamma = 1 - \prod_{j=i-l+1}^{i} (1 - \tilde{Z}_{j,i}^p). \tag{7.5}$$

Since conditions (D1) and (D2) hold, this Z_i^γ-process can be obtained as in Example 4.1, with the duration random variable T and the infinitesimal average patch length L equal to δ_l.

Since the W_i^γ-process now depends on γ, we have to take some care when computing the CVF according to (3.1). However, the explicit formula for CVF given in Theorem 3.1 still holds true with these limiting distributions. Let F_w^γ be the univariate distribution of W_1^γ conditioned on the event $\{Z_1^\gamma = 1\}$ and $F_w^{\gamma,(k)}$ the bivariate distribution of the pair $(W_1^\gamma, W_{1+k}^\gamma)$ conditioned on the event $\{Z_1^\gamma = Z_{1+k}^\gamma = 1\}$. It turns out that F_w^γ and $F_w^{\gamma,(k)}$ converge weakly to some limiting distributions F_w and $F_w^{(k)}$ respectively as $\gamma \to 0$ (When $k \geqq l$, the latter case follows easily from the former, since then $F_w^{\gamma,(k)} = F_w^\gamma \times F_w^\gamma$. However, only $k < l$ is interesting in the limit $\gamma \to 0$.) Introduce $\tilde{V}_j$ as the restriction of V_j^p to the set $\{V_j^p \neq 0\}$. Hence, $\tilde{V}_j$ has distribution H. Let $\mathscr{L}(\cdot)$ and $\mathscr{L}(\cdot,\cdot)$ denote the one- and two-dimensional distributions of the random variables in brackets and $*$ the convolution operator. Finally, put $H_\beta = \mathscr{L}(\beta \tilde{V}_j)$ and $H_{\beta_1,\beta_2} = \mathscr{L}(\beta_1 \tilde{V}_j, \beta_2 \tilde{V}_j)$. We then have:

Theorem 7.1 *Suppose an additive outlier model of the form* (7.1) *for the observed process* $Y_i^\gamma, 0 \leqq \gamma \leqq 1$. *Then* F_w^γ *converges weakly to*

$$F_w = \frac{1}{l} \sum_{j=0}^{l-1} \mathscr{L}(X_1) * \mathscr{L}(\beta_j \tilde{V}_{1-j}) = \frac{1}{l} \sum_{j=0}^{l-1} F_x * H_{\beta_j} \tag{7.6}$$

as $\gamma \to 0$ *and* $F_w^{\gamma,(k)}$, $k=1, \ldots, l-1$ *converge weakly to*

$$F_w^{(k)} = \frac{1}{l-k} \sum_{j=k}^{l-1} \mathscr{L}(X_{1-k}, X_1) * \mathscr{L}(\beta_{j-k}\tilde{V}_{1-j}, \beta_j \tilde{V}_{1-j}) = \frac{1}{l-k} \sum_{j=k}^{l-1} (F_x \times F_x) * H_{\beta_{j-k},\beta_j}. \tag{7.7}$$

Suppose in addition that we have a location M-estimator with a bounded, differentiable score function ψ *whose derivative is also bounded. Then formula* (3.2) *for the* CVF *holds with values of* γ *and* α_k, $k=0, 1, 2, \ldots$ *as specified above and distributions* F_w, $F_w^{(k)}$, $k=1, \ldots, l-1$ *as given in* (7.6)–(7.7).

The forms of $F_w^{(k)}$ in (7.6)–(7.7) reflect the fact that only the event when at most one of l consecutive V_j^p is nonzero is important in the limit $\gamma \to 0$. Note that according to Theorem 7.1, the one-dimensional marginal distributions of $F_w^{(k)}$, $k=1, \ldots, l-1$, do not equal F_w. Therefore, the change-of-variance sensitivity given in Sect. 5 is not applicable in general for the contamination model (7.1). To remedy this, we can take the supremum of CVF$(\psi, \{\mu_y^\gamma\})$ over a larger class $\mathscr{W}$ of W_i-processes with arbitrary marginal distributions except that formula (3.2) remains valid. An upper bound (which is not tight) for this supremum is obtained by bounding each term in the series of formula (3.2) by $\alpha_k \|\psi\|^2/A(\psi) = \alpha_k (\gamma_s^*)^2$ and the sum of the remaining three terms by $\kappa_1^*(\psi, F_x)$. We denote this upper bound by

$$\tilde{\kappa}_\alpha^* = \kappa_1^* + (\alpha - 1)(\gamma_s^*)^2. \tag{7.8}$$

$\tilde{\kappa}_\alpha^*$ is a more general sensitivity than κ_α^*, since it is a bound for a larger class of W-processes. On the other hand, it is a quantity that is more difficult to use when computing optimal or most V_α-robust estimators as in Sect. 6. In principal, one can find the score functions of these estimators by minimizing $V(\psi, \mu_y)$ subject to a *simultaneous* upper bound constraint on κ_1^* and γ_s^* and then minimize these minima over all pairs (κ_1^*, γ_s^*) such that the right-hand side of (7.8) equals a fixed value. The estimators corresponding to the first minimization can be found in about the same way as optimal V-robust estimators are found. In the general case, these estimators are the same as in Subsect. 6.1, whereas for redescending estimators they are different. (Typically, the redescending score function $\psi(\xi)$ will be identical to $A(\xi)$ for small $|\xi|$, tanh-shaped for large $|\xi| (\leq r)$ and constant in a region in between.) The second minimization is more involved, since the minimal value of $V(\psi, \mu_y)$ will depend on (κ_1^*, γ_s^*) in a rather messy way, and it probably has to be carried out numerically.

7.2 Other extensions.

Throughout the paper, we have assumed nominally independent data. Removing this restriction will impose difficulties, since it is then no longer realistic to assume independence between different samples of the X_i- and W_i-processes (condition (C2) of Subsect. 2.3). In order to have a convergent series expressing the asymptotic variance, one has to impose conditions on how fast the dependence between X_i and W_j decreases as $|i-j| \to \infty$.

In order to simplify the exposition, we have confined ourselves to analyse one-dimensional location estimators. By allowing the observed process (2.4) to contain several unknown parameters, one may study the robustness properties of the asymptotic covariance matrix for multidimensional estimators of these parameters. This has been done for independent data by Ronchetti and Rousseeuw (1985).

8 Proofs of theorems

Proof of Theorem 3.1

First some notation: Given an arbitrary process with univariate distribution F and bivariate distributions $F^{(k)}, k=1, 2, \ldots$, let us introduce

$$A(\psi, F)=\int \psi(\xi)^2 \,\mathrm{d} F(\xi), \tag{8.1}$$

$$\tilde{A}(\psi, F^{(k)})=\iint \psi(\xi_1)\, \psi(\xi_2) \,\mathrm{d} F^{(k)}(\xi_1, \xi_2) \tag{8.2}$$

and

$$B(\psi, F)=\int \psi'(\xi) \,\mathrm{d} F(\xi). \tag{8.3}$$

Within regularity, the asymptotic variance is given by (2.3), which in our notation may be written

$$V(\psi, \mu_y^\gamma)=\frac{A(\psi, F_y^\gamma)+2 \sum_{k=1}^{\infty} \tilde{A}(\psi, F_y^{\gamma,(k)})}{B(\psi, F_y^\gamma)} \tag{8.4}$$

for the contaminated process Y_i^γ in (2.4). Each term in the series of the numerator equals

$$\begin{aligned} \tilde{A}(\psi, F_y^{\gamma,(k)}) = & P(Z_1^\gamma=Z_{1+k}^\gamma=0)\, \tilde{A}(\psi, F_x^{(k)}) \\ & +P(Z_1^\gamma=0, Z_{1+k}^\gamma=1)\, \tilde{A}(\psi, F_{xw}^{\gamma,(k)}) \\ & +P(Z_1^\gamma=1, Z_{1+k}^\gamma=0)\, \tilde{A}(\psi, F_{wx}^{\gamma,(k)}) \\ & +P(Z_1^\gamma=Z_{1+k}^\gamma=1)\, \tilde{A}(\psi, F_w^{(k)}). \end{aligned} \tag{8.5}$$

Since ψ is skew-symmetric, f_x is symmetric and the nominal process is i.i.d., the first term in (8.5) vanishes for all $k>0$. Similarly, the second and third terms vanish because of (C2) and the symmetry of f_w. Summing (8.5) over k and using (B3), (C1) and (C4) therefore yields (observe that $A(\psi)=A(\psi, F_x)$):

$$\begin{aligned} & A(\psi, F_y^\gamma)+2 \sum_{k=1}^{\infty} \tilde{A}(\psi, F_y^{\gamma,(k)}) \\ & \quad =(1-g_0(\gamma))\, A(\psi)+g_0(\gamma)\, A(\psi, F_w)+2 \sum_{k=1}^{\infty} g_k(\gamma)\, \tilde{A}(\psi, F_w^{(k)}) \\ & \quad =(1-\gamma)\, A(\psi)+\gamma A(\psi, F_w)+2 \sum_{k=1}^{\infty} \alpha_k\, \gamma \tilde{A}(\psi, F_w^{(k)})+o(\gamma). \end{aligned} \tag{8.6}$$

(The convergence of the series in (8.6) follows from (C4) and dominated convergence.) For the denominator in (8.4) we obtain (observe that $B(\psi)=B(\psi, F_x)$)

$$B(\psi, F_y^\gamma)=(1-g_0(\gamma))\,B(\psi)+g_0(\gamma)\,B(\psi, F_w)$$
$$=(1-\gamma)\,B(\psi)+\gamma B(\psi, F_w)+o(\gamma), \tag{8.7}$$

where the last equality follows from (B4), (2.5) and (C5).

Formula (3.2) now follows by differentiating $\log V(\psi, \mu_y^\gamma)$ with respect to γ, using (8.4), (8.6) and (8.7). □

Proof of Theorem 4.1

Clearly, condition (D2) implies that the Z_i^γ-process is independent of the nominal process. In order to prove the remaining relations, put

$$\gamma=p\sum_{j=0}^{\infty}\theta_j \quad\text{and}\quad \alpha_k=\sum_{j=k}^{\infty}\theta_j\Big/\sum_{j=0}^{\infty}\theta_j, \qquad k=0,1,2,\ldots$$

Then, since $E(T^2)<\infty$ according to (D1), it follows that

$$\sum_{j=0}^{\infty}\theta_j<\infty \quad\text{and}\quad \sum_{k=0}^{\infty}\alpha_k<\infty.$$

It remains to show that $g_k(\gamma)=\alpha_k\gamma+r_k(\gamma)$, with $r_k(\gamma)=o(\gamma)$, $k=0,1,2,\ldots$. According to (4.2), (4.3) and (D2) we have that

$$g_0(\gamma)=P(Z_1^\gamma=1)=1-\prod_{j=0}^{\infty}(1-P(\tilde{Z}_{1-j,1}^p=1))=1-\prod_{j=0}^{\infty}(1-p\theta_j). \tag{8.8}$$

Since the series $\sum_{j=0}^{\infty}p\theta_j$ has nonnegative terms, it holds that

$$1-\sum_{j=0}^{\infty}p\theta_j\leqq\prod_{j=0}^{\infty}(1-p\theta_j)\leqq e^{-\sum_{j=0}^{\infty}p\theta_j} \tag{8.9}$$

and consequently

$$1-e^{-\gamma}\leqq g_0(\gamma)\leqq\gamma, \tag{8.10}$$

which yields the desired conclusion for $k=0$. For $k>0$, put

$$D_1^{p,k}=\bigcup_{j=0}^{\infty}\{\tilde{Z}_{1-j,1+k}^p=1\}, \tag{8.11}$$

$$D_2^{p,k}=\bigcup_{j=0}^{\infty}\{\tilde{Z}_{1-j,1}^p=1\}\setminus D_1^{p,k} \tag{8.12}$$

and

$$D_3^{p,k}=\bigcup_{j=0}^{k-1}\{\tilde{Z}_{1+k-j,1+k}^p=1\}, \tag{8.13}$$

so that

$$\{Z_1^\gamma=Z_{1+k}^\gamma=1\}=D_1^{p,k}\cup(D_2^{p,k}\cap D_3^{p,k}\}. \tag{8.14}$$

By similar arguments as for $k=0$ we obtain

(8.15) $$P(D_1^{p,k})=1-\prod_{j=k}^{\infty}(1-p\theta_j)=p\sum_{j=k}^{\infty}\theta_j+o(\gamma)=\gamma\alpha_k+o(\gamma),$$

(8.16) $$P(D_2^{p,k})\leqq 1-\prod_{j=0}^{\infty}(1-p\theta_j)\leqq\gamma$$

and

(8.17) $$P(D_3^{p,k})=1-\prod_{j=0}^{k-1}(1-p\theta_j)\leqq p\sum_{j=0}^{k-1}\theta_j\leqq\gamma.$$

$D_1^{p,k}$ and $D_2^{p,k}\cap D_3^{p,k}$ are disjoint events and by (D2), $D_2^{p,k}$ and $D_3^{p,k}$ are independent. Hence,

(8.18) $$g_k(\gamma)=P(D_1^{p,k})+P(D_2^{p,k})\,P(D_3^{p,k})=\alpha_k\,\gamma+o(\gamma). \quad \square$$

Proof of Theorem 4.2

Because of stationarity, the distribution of $\tilde{L}_i^\gamma$ is clearly independent of i, and hence we may assume $i=0$. Notice that if $Z_0^\gamma=1$,

(8.19) $$\tilde{L}_0^\gamma=\tilde{L}_0^{\gamma,+}-\tilde{L}_0^{\gamma,-}+1,$$

where

$$\tilde{L}_0^{\gamma,+}=\max\{j\geqq 0;\, Z_0^\gamma=\ldots=Z_j^\gamma=1\}$$

and

$$\tilde{L}_0^{\gamma,-}=\min\{j\leqq 0;\, Z_j^\gamma=\ldots=Z_0^\gamma=1\}.$$

Again, because of stationarity,

$$P(\tilde{L}_0^{\gamma,+}=k,\tilde{L}_0^{\gamma,-}=l)=q_{k-l},\ k\geqq 0,\ l\leqq 0,$$

and hence by (8.19),

(8.20) $$P(\tilde{L}_0^\gamma=j)=j\,q_{j-1}.$$

In order to calculate q_j, observe that

(8.21) $$\begin{aligned} q_j&=P(\tilde{L}_0^{\gamma,-}=0,\tilde{L}_0^{\gamma,+}=j)\\ &=P(Z_0^\gamma=\ldots=Z_j^\gamma=1,Z_{-1}^\gamma=Z_{j+1}^\gamma=0)\\ &=P\left(\bigcap_{k\leqq -1}\{\tilde{Z}_{k,-1}^\gamma=0\},\bigcap_{k=0}^{j+1}\{\tilde{Z}_{k,j+1}^\gamma=0\},\{\tilde{Z}_{0,j}^\gamma=1\}\right)+o(\gamma)\\ &=\frac{\gamma P(T=j+1)}{E(T)}+o(\gamma),\end{aligned}$$

where the last two equalities follows by similar (but slightly more complicated) arguments as in the proof of Theorem 4.1. It then follows from (8.20)–(8.21) that

$$P(L_0^\gamma=j)=\frac{P(\tilde{L}_0^\gamma=j)}{g_0(\gamma)}=\frac{jP(T=j)}{E(T)}+o(1), \quad j=1,2,\ldots,$$

which proves that $L_0^\gamma \xrightarrow{d} L$ as $\gamma\to 0$. In order to show that $L \overset{s}{\geqq} T$, it follows from (4.4)–(4.5), (4.8) and partial summation that

$$(8.22)\quad P(L>j)=\frac{(j+1)\,\theta_j+\alpha_{j+1}\,E(T)}{\sum_{k=0}^{j}\theta_k+\alpha_{j+1}\,E(T)} \geqq \frac{(j+1)\,\theta_j+\alpha_{j+1}\,E(T)}{j+1+\alpha_{j+1}\,E(T)} \geqq \theta_j=P(T>j),$$

with equality iff $\theta_j=0$ or 1. In order for T and L to have the same distribution, θ_j must equal 0 or 1 for all j, and hence $T=\delta_l$ for some $l\geqq 1$. Finally, (4.9) follows easily from the definitions of L and T. □

Proof of Theorem 6.1

The first part of the theorem (V_α-robustness versus B-robustness) is proved in the same way as for the case $\alpha=1$ (Hampel et al. 1986, Theorems 2.5.1–2.5.2). For the second part, the first inequality in (6.2) is obvious while the second one follows, since $\kappa_\alpha^*(\psi,F_x)\leqq\kappa_1^*(\psi,F_x)+(\alpha-1)\,\gamma_s^*(\psi,F_x)^2\leqq\kappa_1^*(\psi,F_x)+(\alpha-1)(\kappa_1^*(\psi,F_x)-1)$ by (6.1). Finally, it follows that the last inequality of (6.2) can be replaced by equality for nondecreasing ψ, since (6.1) then holds with equality for both $\kappa_1^*(\psi,F_x)$ and $\kappa_\alpha^*(\psi,F_x)$. □

Proof of Theorem 7.1

We start by proving weak convergence of F_w^γ and $F_w^{\gamma,(k)}$, $k=1,\ldots,l-1$. Actually, we will prove the stronger relations

$$(8.23)\qquad F_w^\gamma=(1-\varepsilon_0(\gamma))\,F_w+\varepsilon_0(\gamma)\,\tilde{F}_w^\gamma$$

and

$$(8.24)\qquad F_w^{\gamma,(k)}=(1-\varepsilon_k(\gamma))\,F_w^{(k)}+\varepsilon_k(\gamma)\,\tilde{F}_w^{\gamma,(k)}, \qquad k=1,\ldots,l-1,$$

where $\varepsilon_k(\gamma)=o(1)$, $k=0,\ldots,l-1$ and $\tilde{F}_w^\gamma$ and $\tilde{F}_w^{\gamma,(k)}$, $k=1,\ldots,l-1$ are distributions that will be specified below.
Define

$$(8.25)\qquad E_k^p=\{V_{1-k}^p\neq 0\}\cap\left(\bigcap_{\substack{j=0\\ j\neq k}}^{l-1}\{V_{1-j}^p=0\}\right),$$

and let $\tilde{E}^p$ be the event that at least two of the random variables $V_1^p,\ldots,V_{2-l}^p$ are nonzero. Then according to (7.5) we have

$$(8.26)\qquad \{Z_1^\gamma=1\}=\left(\bigcup_{k=0}^{l-1}E_k^p\right)\cup\tilde{E}^p.$$

Since the V_j^p's are independent, (8.25) gives

$$P(E_k^p)=p(1-p)^{l-1} \tag{8.27}$$

when $0 \leqq k \leqq l-1$ and consequently (remember that $\gamma = lp$)

$$p(1-\gamma) \leqq P(E_k^p) \leqq p. \tag{8.28}$$

The events E_k^p, $k=0, \ldots, l-1$ and $\tilde{E}^p$ are clearly disjoint. (2.5) and (8.28) therefore imply

$$P(\tilde{E}^p)=g_0(\gamma)-\sum_{k=0}^{l-1} P(E_k^p)=o(\gamma). \tag{8.29}$$

Since V_i^p is an i.i.d. process, independent of the nominal process, the conditional distribution of W_1^γ under E_k^p is $F_x * H_{\beta_k}$. Let $\tilde{F}_w^\gamma$ be the conditional distribution of W_1^γ under $\tilde{E}^p$. Then we have

$$F_w^\gamma=\sum_{k=0}^{l-1} \frac{P(E_k^p)}{g_0(\gamma)} F_x * H_{\beta_k}+\frac{P(\tilde{E}^p)}{g_0(\gamma)} \tilde{F}_w^\gamma, \tag{8.30}$$

which implies (8.23). In order to prove (8.24), we proceed in an analogous manner, using the partition

$$\{Z_1^\gamma=Z_{1-k}^\gamma=1\}=\left(\bigcup_{j=k}^{l-1} E_j^p\right) \cup \tilde{E}_k^p, \tag{8.31}$$

where the "remaining" event $\tilde{E}_k^p$ is negligible in the limit $\gamma \to 0$. The bivariate distribution of $(W_{1-k}^\gamma, W_1^\gamma)$, conditioned on the event E_j^p, equals $\mathscr{L}(X_{1-k}, X_1) * \mathscr{L}(\beta_{j-k} \tilde{V}_{1-j}, \beta_j \tilde{V}_{1-j})=(F_x \times F_x) * H_{\beta_{j-k}, \beta_j}$ when $k \leqq j \leqq l-1$. Letting $\tilde{F}_w^{\gamma,(k)}$ be the conditional distribution of $(W_{1-k}^\gamma, W_1^\gamma)$ under $\tilde{E}_k^p$, we obtain from (8.31)

$$F_w^{\gamma,(k)}=\sum_{j=k}^{l-1} \frac{P(E_j^p)}{g_k(\gamma)} (F_x \times F_x) * H_{\beta_{j-k}, \beta_j}+\frac{P(\tilde{E}_k^p)}{g_k(\gamma)} \tilde{F}_w^{\gamma,(k)}, \tag{8.32}$$

which implies (8.24).

We proceed by proving the validity of formula (3.2) for the CVF. Remembering the notation in the proof of Theorem 3.1, the asymptotic variance is given by (8.4). Each term in the series of that formula can be expanded as

$$\begin{aligned} \tilde{A}(\psi, F_y^{\gamma,(k)}) = & P(Z_1^\gamma=Z_{1+k}^\gamma=0)\, \tilde{A}(\psi, F_x^{(k)}) \\ & +P(Z_1^\gamma=0, Z_{1+k}^\gamma=1)\, \tilde{A}(\psi, F_{xw}^{\gamma,(k)}) \\ & +P(Z_1^\gamma=1, Z_{1+k}^\gamma=0)\, \tilde{A}(\psi, F_{wx}^{\gamma,(k)}) \\ & +P(Z_1^\gamma=Z_{1+k}^\gamma=1)\, \tilde{A}(\psi, F_w^{\gamma,(k)}). \end{aligned} \tag{8.33}$$

The first three terms in (8.33) vanish for the same reason as in Theorem 3.1. (Condition (C2) of Subsect. 2.3 is obviously satisfied for the process (7.1).) Summing over k therefore yields (the extra γ's in the superscripts compared to (8.5) indicate that the W_i^γ-process now depends on γ)

$$(8.34)\qquad \begin{aligned} &A(\psi, F_y^\gamma)+2\sum_{k=1}^{\infty} \tilde{A}(\psi, F_y^{\gamma,(k)}) \\ &\quad=(1-g_0(\gamma))\,A(\psi)+g_0(\gamma)\,A(\psi, F_w^\gamma)+2\sum_{k=1}^{\infty} g_k(\gamma)\,\tilde{A}(\psi, F_w^{\gamma,(k)}) \\ &\quad=(1-\gamma)\,A(\psi)+\gamma A(\psi, F_w)+2\sum_{k=1}^{\infty} g_k(\gamma)\,\tilde{A}(\psi, F_w^{\gamma,(k)})+o(\gamma), \end{aligned}$$

where the last equality follows from (2.5), the fact that $A(\psi)$ and $A(\psi, F_w^\gamma)$ are bounded (since ψ is bounded) and from (8.23) (which implies $A(\psi, F_w^\gamma) \to A(\psi, F_w)$ as $\gamma \to 0$). Next we show that

$$(8.35)\qquad \sum_{k=1}^{\infty} g_k(\gamma)\,\tilde{A}(\psi, F_w^{\gamma,(k)})=\sum_{k=1}^{\infty} \alpha_k\,\gamma \tilde{A}(\psi, F_w^{(k)})+o(\gamma).$$

When $k \geqq l$, W_1^γ and W_{1+k}^γ are conditionally independent under the event $\{Z_1^\gamma = Z_{1+k}^\gamma = 1\}$ and hence $\tilde{A}(\psi, F_w^{\gamma,(k)})=0$. When $k<l$, the boundedness of ψ together with (8.24) imply that $\tilde{A}(\psi, F_w^{\gamma,(k)}) \to \tilde{A}(\psi, F_w^{(k)})$ as $\gamma \to 0$. Formula (8.35) then follows by dominated convergence. The denominator of (8.4) may be written

$$(8.36)\qquad \begin{aligned} B(\psi, F_y^\gamma) &= (1-g_0(\gamma)\,B(\psi)+g_0(\gamma)\,B(\psi, F_w^\gamma) \\ &= (1-\gamma)\,B(\psi)+\gamma B(\psi, F_w)+o(\gamma). \end{aligned}$$

The last equality follows since $B(\psi, F_w^\gamma) \leqq \|\psi'\|$ uniformly in γ and since $B(\psi, F_w^\gamma) \to B(\psi, F_w)$ as $\gamma \to 0$, which in turn is a consequence of (8.23) and the boundedness of ψ'. The desired conclusion now follows by differentiating $\log V(\psi, \mu_y^\gamma)$ with respect to γ, using (8.4), (8.34), (8.35) and (8.36). □

Acknowledgement. This paper is part of my PhD thesis, and I am grateful to my advisor Professor Svante Janson for valuable comments on the work. I also thank the associate editor and a referee for their suggestions, especially regarding the interpretation of α.

References

Beran, J., Künsch, H.: Location estimators for processes with long range dependence. Research report No 40, Sem. für Statist. ETH, 1985

Billingsley, P.: Convergence of probability measures. New York: Wiley 1968

Bustos, O.H.: General M-estimates for contaminated pth-order autoregressive processes: Consistency and asymptotic normality. Z. Wahrscheinlichkeitstheor. Verw. Geb. **59**, 491–504 (1982)

Fox, R., Taqqu, M.S.: Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. Ann. Stat. **14**, 517–532 (1986)

Franke, J., Hannan, E.J.: Comment on "Influence functionals for time series". Ann. Stat. **14**, 822–824 (1986)

Hampel, F.R.: The influence curve and its role in robust estimation. J. Am. Stat. Assoc. **69**, 383–393 (1974)

Huber, P.J.: Robust estimation of a location parameter. Ann. Math. Stat. **35**, 73–101 (1964)

Huber, P.J.: The behaviour of maximum likelihood estimates under nonstandard conditions. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 221–233, 1967

Hössjer, O.: The change-of-variance function for dependent data. U.U.D.M. Report 1989:18, Dept. of Mathematics, Uppsala University, 1989

Künsch, H.: Infinitesimal robustness for autoregressive processes. Ann. Stat. **12**, 843–863 (1984)

Künsch, H.: Statistical aspects of self-similar processes. Bernoulli **1**, 67–74 (1987)

Lee, C.-H., Martin, R.D.: Ordinary and proper M-estimates for autoregressive-moving average models. Biometrika **73**, 679–686 (1986)

Mandelbrot, B.B.: The fractal geometry of nature. New York: Freeman 1977

Martin, R.D., Yohai, V.J.: Influence functionals for time series. Ann. Stat. **14**, 781–818 (1986a)

Martin, R.D., Yohai, V.J.: Reply to "Influence functionals for time series". Ann. Stat. **14**, 840–855 (1986b)

Moustakides, G.V., Thomas, J.B.: Min-max detection of weak signals in ϕ-mixing noise. IEEE Trans. Inf. Theory **30**, 529–537 (1984)

Noether, G.E.: On a theorem of Pitman. Ann. Math. Stat. **26**, 64–68 (1955)

Portnoy, S.L.: Robust estimation in dependent situations. Ann. Stat. **5**, 22–43 (1977)

Portnoy, S.L.: Further remarks on robust estimation in dependent situations. Ann. Stat. **7**, 224–231 (1979)

Ronchetti, E., Rousseeuw, P.J.: Change-of-variance sensitivities in regression analysis. Z. Wahrscheinlichkeitstheor. Verw. Geb. **68**, 503–519 (1985)

Rousseeuw, P.J.: A new infinitesimal approach to robust estimation. Z. Wahrscheinlichkeitstheor. Verw. Geb. **56**, 127–132 (1981)

Rousseeuw, P.J.: Most robust M-estimators in the infinitesimal sense. Z. Wahrscheinlichkeitstheor. Verw. Geb. **61**, 541–555 (1982)

Sadowsky, J.S.: A maximum variance model for robust detection and estimation with dependent data. IEEE Trans. Inf. Theory **32**, 220–226 (1986)

Zamar, H.: Robustness against unexpected dependence in the location model. Stat. Probab. Lett. **9**, 367–374 (1990)