The Annals of Statistics
1977, Vol. 5, No. 1, 22-43

ROBUST ESTIMATION IN DEPENDENT SITUATIONS[1]

BY STEPHEN L. PORTNOY

University of Illinois, Urbana-Champaign

To analyze the effect of correlation in random samples on the performance of estimators of location, small correlation approximations for the asymptotic variance are found. Approximately optimal estimators (in the asymptotic minimax sense of Huber) are presented and compared to other estimators in terms of maximum asymptotic variance over the class of ε-contaminated normals. The presence of relatively small correlation can drastically inflate variances, and the optimal rules given here offer substantial improvements over previously considered estimators.

1. Introduction and optimal asymptotic variance. The effect of dependence on robust procedures has previously been studied only for particular procedures and particular distributional situations (e.g., see Gastwirth and Rubin [3]). However, the problem of finding optimal robust procedures has not been solved. In this paper, estimators of a location parameter which are approximately asymptotically optimal in the sense of Huber ([4], [5]) are found for models which are similar to (and asymptotically equivalent to) moving average schemes. The basic model is as follows: let $Y_1, Y_2, \cdots, Y_n$ be i.i.d. with continuous, symmetric cdf G, let $Y_0 = Y_n$ and $Y_{n+1} = Y_1$, and define for $i = 1, 2, \cdots, n$

$$X_i = \theta + Y_i + \rho Y_{i-1} + \rho Y_{i+1} \tag{1.1}$$

where θ is a location parameter and ρ is a parameter with $|\rho| < 1$ (and which will generally be taken to tend to zero). The reason for defining Y_0 and Y_{n+1} as above is so that the distribution of $X_1, \cdots, X_n$ will be stationary (and, hence, the marginal distributions will be the same), and so that the result of Theorem 1.1 (providing the optimal asymptotic variance) will apply directly.

Section 2 presents the basic expansion of the asymptotic variance for rather general estimators in the form

$$\sigma^2 = \sigma_0^2 + c\rho + \mathcal{O}(\rho^2) \tag{1.2}$$

where σ_0^2 would be the variance for an independent sample and c depends on the distribution, G. Section 3 uses methods of Huber ([4] and [5]) to find estimators minimizing the maximum approximate asymptotic variance (1.2) over distributions, G, which are ε-contaminated normals. The minimax estimator corresponds to a maximum likelihood-type estimator (M-estimator) defined by

Received February 1975; revised May 1976.

[1] This work was facilitated by National Science Foundation Grant GP-31003X and MPS 75-07978.

AMS 1970 *subject classifications.* Primary 62G35; Secondary 62F10.

Key words and phrases. Robust estimation, dependent observations, asymptotic minimaxity over ε-contaminated normals.

a function Ψ which is piecewise linear with slope one near the origin and slope $-\rho$ in the tails. Thus, the possible presence of positive correlation can justify the use of Ψ-functions with a small amount of redescent.

Section 4 discusses how to find the maximum approximate asymptotic variance, σ^2, over the class of ε-contaminates of $\mathcal{N}(0, 1)$, and lists numerical results for a number of estimators. Conclusions are presented at the end of the section.

Section 5 presents an expansion of σ^2 to terms in ρ^3 and uses it to assess the adequacy of the first term approximation. The appendix presents a theorem providing asymptotic normality for general classes of M-estimators in m-dependent situations. Since previously published theorems apply only to continuously differentiable Ψ functions (even in the independent case) the theorem here (which permits the derivative to be discontinuous) should be of independent interest.

Now consider a generalization of (1.1): let $\mathbf{Y} = (Y_1, \cdots, Y_n)'$ and assume the observation vector $\mathbf{X} = (X_1, \cdots, X_n)'$ satisfies

$$\mathbf{X} = \theta\mathbf{e} + S\mathbf{Y} \tag{1.3}$$

where $\mathbf{e} = (1, \cdots, 1)'$ and S is an $n \times n$ symmetric circulant matrix with first row $(1\ \rho_1\ \rho_2\ \rho_3 \cdots \rho_2\ \rho_1)$ (such that S is invertible). If S is known, this model can be easily reduced to the independent case. Let $d = 1 + \rho_1 + \rho_2 + \rho_3 + \cdots + \rho_2 + \rho_1$. Then $S\mathbf{e} = d\mathbf{e}$, so $S^{-1}\mathbf{e} = (1/d)\mathbf{e}$; and, hence,

$$S^{-1}\mathbf{X} = \theta S^{-1}\mathbf{e} + \mathbf{Y} = (\theta/d)\mathbf{e} + \mathbf{Y}\,.$$

Therefore,

$$dS^{-1}\mathbf{X} = \theta\mathbf{e} + d\mathbf{Y}\,. \tag{1.4}$$

Thus, given an optimal estimator in the independent case, that procedure can be directly applied to the coordinates of $dS^{-1}\mathbf{X}$ to obtain an optimal estimator for model (1.3). The major difference is that the variance will be inflated by an amount d^2; and, hence, it is clear that relatively small amounts of dependence spread throughout a sample can rather seriously inflate the variance of estimators over what would be expected in an independent sample.

In particular, combining these results with known results on efficiency of estimators will provide lower bounds for the asymptotic variance of any sequence of estimators. The recent results of Stone [7] will therefore yield

THEOREM 1.1. *Let σ_0^2 be the inverse Fisher information for the distribution of Y as defined by Stone* [9]. *Let $\{\tilde{\theta}_n\}$ be a sequence of estimators such that $n^{\frac{1}{2}}(\hat{\theta}_n - \theta) \rightarrow_D \mathcal{N}(0, \sigma^2(\theta))$. Then* (i) *$\sigma^2(\theta) \geqq d^2\sigma_0^2$ for almost all θ, and* (ii) *if $\{\hat{\theta}_n\}$ are location invariant then $\sigma^2(\theta) \geqq d^2\sigma_0^2$ for all θ.*

We are interested in considering the effect of dependence on various estimators based on the original sample $(X_1, X_2, \cdots, X_n)$. Since the relationship between $\mathbf{X}$ and $dS^{-1}\mathbf{X}$ is generally difficult to consider, the remainder of this paper is concerned mainly with the special case of (1.1). However, it should be possible to extend most results (at least qualitatively) to model (1.3).

2. The basic small-correlation expansion of the asymptotic variance. Let $(X_1, X_2, \cdots, X_n)$ be a random vector with stationary marginal distribution F, where (setting the parameter $\theta = 0$) F is a continuous cdf symmetric about zero. Let T be a real valued functional on the space of all cdf's, and let F_n be the empirical cdf of the sample. Then Filippova [2] shows (under conditions in the independent case) that

$$T(F_n) = T(F) + \int I_F(x)\, dF_n(x) + o_p(1/n^{\frac{1}{2}}) \tag{2.1}$$

as $n \to \infty$, where $I_F(x)$ is the influence curve (or von Mises derivative) of T at F:

$$I_F(x) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \{T((1-\varepsilon)F + \varepsilon\delta_x) - T(F)\} \tag{2.2}$$

(and where δ_x denotes a unit point mass at x). Assume $I_F(x)$ is skew-symmetric about zero, and that $T(F) = 0$.

The main stochastic term is

$$\int I_F(x)\, dF_n(x) = \frac{1}{n} \sum_{i=1}^n I_F(X_i) ,$$

so that as long as the dependence is such that the central limit theorem holds, we may expect

$$n^{\frac{1}{2}} T(F_n) \to_D \mathcal{N}(0, \sigma^2) \tag{2.3}$$

where

$$\begin{aligned} \sigma^2 &= EI_F^2(X) + 2 \sum_{i=2}^\infty \operatorname{Cov}(I_F(X_1), I_F(X_i)) \\ &= EI_F^2(X) + 2 \sum_{i=2}^\infty EI_F(X_1) I_F(X_i) . \end{aligned} \tag{2.4}$$

It should be possible to prove a version of Filippova's result (2.1) (which requires essentially only the weak law of large numbers). However, the conditions would be at least as complicated as those in [2]. The appendix proves (2.3) for most M estimators when the observations are m-dependent. Formally, an M-estimator is given by defining $T(F)$ to be any value satisfying

$$\int \Psi(x - T(F))\, dF(x) = 0$$

where here Ψ will be an odd real valued function. The fact that (2.4) agrees with the expression in the appendix follows from the fact that the influence curve for an M-estimator (as given in [5]) is

$$I_G(x) = \Psi(x)/E\Psi'(Y) \tag{2.5}$$

if Ψ is absolutely continuous (and where expectation is with respect to G).

The following result now evaluates σ^2 in (2.4).

THEOREM 2.1. *Under model* (1.1) *assume that* F (*the marginal of* X_i) *and* T *are such that*

$$|I_F(x)| \leqq C_1 \qquad \text{and} \qquad \|I_F(x) - I_G(x)\| \leqq C_2 \|F - G\| , \tag{2.6}$$

and that G has a finite second moment and has a characteristic function, $\varphi_Y(u)$, such that $\int u^2|\varphi_Y(u)|\,du < +\infty$. (This implies that G has a density, g, with a continuous bounded second derivative.) Then, with σ^2 given by (2.4),

$$\sigma^2 = EI_G^2(Y) - 4\rho EYI_G(Y) \int I_G(x)g'(x)\,dx + \mathcal{O}(\rho^2)\,. \tag{2.7}$$

REMARK. If $I_G(x)$ is absolutely continuous, (2.7) implies

$$\sigma^2 = EI_G^2(Y) + 4\rho EYI_G(Y)EI_G'(Y) + \mathcal{O}(\rho^2)\,. \tag{2.8}$$

Actually the regularity conditions on G can be transferred to $I_G(x)$; so that if $EY^2 < +\infty$, I_G has a bounded, continuous second derivative and the second part of (2.6) holds; then expanding $I_G(x)$ in a Taylor series immediately yields (2.8) (even if $I_F(x)$ is not bounded).

PROOF OF THEOREM. Under model (1.1), two observations are independent if their indices differ by more than two; so only the variance and first two covariance terms are needed to apply (2.4). Consider the characteristic functions of X and of the pairs (X_1, X_2) and (X_1, X_3). Since $EY^2 < +\infty$, these can be expanded using

$$\exp\{iu[Y_1 + \rho(Y_2 + Y_n)]\} = e^{iuY_1}[1 + iu\rho(Y_2 + Y_n) + u^2(Y_2^2 + Y_n^2)\mathcal{O}(\rho^2)]\,.$$

This yields (since $EY = 0$)

$$\varphi_X(u) = \varphi_Y(u)[1 + u^2\mathcal{O}(\rho^2)]\,, \tag{2.9}$$

which can be inverted to obtain

$$f_X(x) = g(x) + g''(x)\mathcal{O}(\rho^2)\,. \tag{2.10}$$

Hence, using (2.6), $EI_F^2(X) = EI_G^2(Y) + \mathcal{O}(\rho^2)$. Similarly

$$\begin{aligned}\varphi_{(X_1,X_2)}(u, v) &= \varphi_Y(u + \rho v)\varphi_Y(v + \rho u)[1 + \mathcal{O}(\rho^2)] \\ &= [\varphi_Y(u)\varphi_Y(v) + \rho\{u\varphi_Y(u)\varphi_Y'(v) + v\varphi_Y(v)\varphi_Y'(u)\} \\ &\qquad + \rho^2\tilde{\varphi}(u, v, \rho)][1 + \mathcal{O}(\rho^2)]\end{aligned}$$

where $\tilde{\varphi}(u, v, \rho) \to -\frac{1}{2}[u^2\varphi_Y(u)\varphi_Y''(v) + 2uv\varphi_Y'(u)\varphi_Y'(v) + v^2\varphi_Y(v)\varphi_Y''(u)]$ as $\rho \to 0$ for each u and v. By hypothesis, $g''(x)$ is continuous and uniformly bounded, so by change of variables and direct expansion,

$$f_{(Y_1+\rho Y_2, Y_2+\rho Y_1)}(x, y) = g(x)g(y) - \rho\{yg(y)g'(x) + xg(x)g'(y)\} + \rho^2\tilde{f}(x, y, \rho)$$

where $\tilde{f}(x, y, \rho)$ is uniformly bounded in ρ. Comparing the above expansions, $\tilde{\varphi}(u, v, \rho)$ is the Fourier transform of $\tilde{f}(x, y, \rho)$. Thus, using basic properties of the convergence of Fourier transforms, for any bounded measurable function $h(x, y)$,

$$\begin{aligned}\iint h(x, y)\tilde{f}(x, y, \rho)\,dx\,dy \to \iint h(x, y)[&x^2g(x)g''(y) + 2g'(x)g'(y) \\ &+ y^2g(y)g''(x)]\,dx\,dy\end{aligned}$$

as $\rho \to 0$. Therefore,

$$EI_F(X_1)I_F(X_2) = -2\rho EYI_F(Y) \int I_F(x)g'(x)\,dx + \mathcal{O}(\rho^2)\,.$$

In a similar manner,

$$EI_F(X_1)I_F(X_3) = \mathcal{O}(\rho^2)\,.$$

The theorem follows using the last part of (2.6) and (2.10).

For M-estimators, if Ψ is absolutely continuous with Ψ' bounded, then (2.10) implies that (2.6) holds. Thus, in this case, (2.8) becomes

$$\sigma^2 = \frac{E\Psi^2(Y)}{(E\Psi'(Y))^2} + 4\rho \frac{EY\Psi(Y)}{E\Psi'(Y)} + \mathcal{O}(\rho^2) . \tag{2.11}$$

In other types of estimators, the influence curve, I_F will generally depend more explicitly on F, and the resulting formulas for σ^2 will tend to be substantially more complicated. However, results for a restricted class of estimators which are linear combinations of order statistics (including trimmed means) can be easily found. Particular examples are presented in Sections 3 and 4.

3. Asymptotic minimaxity. This section considers the problem of minimizing the maximum approximate asymptotic variance over various classes of distributions, G, in model (1.1). Since any influence curve can be achieved by an M-estimator with appropriate ψ function, we will restrict attention to M-estimators and use formula (2.11) which gives the approximate asymptotic variance under model (1.1) with $Y \sim G$. Also assume throughout this section that G has a density g, and begin heuristically by considering the approximate asymptotic variance in (2.11) here denoted by $V_1(g, \psi)$.

We first fix g and find the function ψ_1 (as a function of g) which minimizes $V_1(g, \psi)$. We then show that the minimum asymptotic variance $V_1^*(g) = V_1(g, \psi^*)$ is simply $(1 + 4\rho)V_0^*(g) + \mathcal{O}(\rho^2)$ where

$$V_0^*(g) = V_0(g, \psi_0) = \frac{E\psi_0^2(Y)}{(E\psi_0'(Y))^2} \tag{3.1}$$

is the minimum asymptotic variance in the independent case. It follows that the least favorable distribution for $V_1(g, \psi)$ over any class of distributions is the same as that for $V_0(g)$. Since Huber ([4] and [5]) has found this least favorable distribution in certain cases, his results can be directly applied to find a minimax procedure for the asymptotic variance V_1 (under model (1.1)). This will be done explicitly for the class of ε-contaminates of an $\mathcal{N}(0, 1)$ distribution, where a formal optimality theorem will be proven. The equivalent optimal linear combination of order statistics will also be given.

From results of Huber, $V_0(g, \psi)$ is convex in ψ (for fixed g) and will have a unique minimum, ψ_0, over the class of ψ for which $V_0(g, \psi)$ is finite. Hence, for ρ small, $V_1(g, \psi)$ will also have a unique minimum, ψ_1, in a neighborhood of ψ_0. Using standard variational methods, if ψ_1 is a unique minimum for V_1, then the directional derivative in direction h equals zero for all functions h. In particular, we want to find ψ_1 which satisfies the following equation for any continuously differentiable function h vanishing outside a compact set:

$$\begin{aligned} 0 &= \lim_{t\to\infty} \frac{1}{t} \{V_1(g, \psi + th) - V_1(g, \psi)\} \\ &= \frac{2E\psi(Y)h(Y)}{(E\psi'(Y))^2} - \frac{2E\psi^2(Y)Eh'(Y)}{(E\psi'(Y))^3} + 4\rho \frac{EYh(Y)}{E\psi'(Y)} - 4\rho \frac{EY\psi(Y)Eh'(Y)}{(E\psi'(Y))^2} \end{aligned} \tag{3.2}$$

where Y has density g. Thus, letting

$$a = E\psi'(Y)\,, \qquad b = E\psi^2(Y) \qquad \text{and} \qquad c = EY\psi(Y)\,, \tag{3.3}$$

the minimizing ψ_1 must satisfy

$$0 = 2aE\psi(Y)h(Y) - 2bEh'(Y) + 4\rho a^2 EYh(Y) - 4\rho acEh'(Y)\,.$$

Now integrating $Eh'(Y)$ by parts yields

$$0 = \int [2a\psi(y)g(y) + 2bg'(y) + 4\rho a^2 yg(y) + 4\rho acg'(y)]h(y)\,dy\,. \tag{3.4}$$

This implies that (at least for almost every y) the integrand in brackets in (3.4) must vanish. Hence solving for the minimizing ψ_1,

$$\begin{aligned}\psi_1(x) &= -2\rho ax - \left(\frac{2b + 4\rho ac}{2a}\right)\frac{g'(x)}{g(x)} \\ &= -\frac{b}{a}\left[\left(1 + \frac{2ac}{b}\rho l'(x)\right) + \frac{2a^2}{b}\rho x\right]\end{aligned} \tag{3.5}$$

where

$$l'(x) = \frac{d}{dx}\log g(x) = \frac{g'(x)}{g(x)}\,.$$

Since M-estimators are unchanged if ψ is multiplied by a constant, we may take the minimizing ψ_1 to be

$$\psi_1(x) = -(1 + \beta\rho)l'(x) - \rho\alpha x \tag{3.6}$$

where (with a, b and c given in (3.3)),

$$\alpha = \frac{2a^2}{b} \qquad \text{and} \qquad \beta = \frac{2ac}{b}\,. \tag{3.7}$$

Note that the function ψ_0 minimizing $V_0(g, \psi)$ is $\psi_0(x) = -l'(x)$. Hence,

$$\psi_1(x) = \psi_0(x) - \rho(\alpha x + \beta l'(x))\,,$$

and the first order correction to ψ_0 in the dependent model is to subtract the odd function $(\alpha x + \beta l'(x))$.

From Theorem 1.1 with $d = (1 + 2\rho)$ the minimum asymptotic variance must be

$$V_1^*(g) = (1 + 2\rho)^2 V_0^*(g) \doteq (1 + 4\rho + \mathcal{O}(\rho^2))V_0^*(g) \tag{3.8}$$

where $V_0^*(g)$ is given by (3.1). This could be obtained directly by inserting (3.5) in $V_1(g, \psi_1)$ and using first order approximations. Thus, for any family of distributions the least favorable distribution is the same as in the independent case. In particular, for the family $\mathscr{F}_1$ of ε-contaminates of $\mathscr{N}(0, 1)$, the least favorable distribution (as found in Huber [4]) has

$$\begin{aligned}-l'(x) = -\frac{d}{dx}\log g(x) &= x & |x| < k \\ &= k\,\mathrm{sgn}\,x\,, & |x| > k\end{aligned}$$

where k depends on ε and satisfies

$$2\Phi(k) - 1 + \frac{2}{k}\phi(k) = \frac{1}{1-\varepsilon}, \tag{3.9}$$

where $\Phi(x)$ is the unit normal cdf and $\phi(x)$ is the unit normal density. Hence, to apply (3.6) to find the minimax rule, it remains to compute α and β, at least to first order terms. To do this requires computing a, b and c (3.4) for $\psi(x) = \psi_0(x)$; and this only requires $El''(Y)$, $E(l'(Y))^2$, and $EYl'(Y)$, where l is the logarithmic derivative of the least favorable density and Y has this least favorable distribution. Direct computations (which use (3.9)) yield

$$\alpha = 2(1-\varepsilon)(2\Phi(k)-1) \qquad \text{and} \qquad \beta = 2\,.$$

Therefore, formula (3.10) yields the function ψ_1, which is minimax over the class $\mathscr{F}_1$ with respect to the approximate asymptotic variance $V_1(g, \psi)$:

$$\begin{aligned}
\psi_1(x) &\propto \begin{cases} c[1+\rho(\beta-\alpha)]x & |x| \leqq k \\ c(1+\rho\beta)k \operatorname{sgn} x - \alpha\rho x & |x| > k \end{cases} \\
&\propto \begin{cases} x & |x| \leqq k \\ k \operatorname{sgn} x - \dfrac{\rho\alpha}{1+\rho(\beta-\alpha)}(x - k \operatorname{sgn} k) & |x| > k \end{cases} \\
&= \begin{cases} x & |x| \leqq k \\ k \operatorname{sgn} x - \dfrac{2\rho a}{1+2\rho(1-a)}(x - k \operatorname{sgn} x) & |x| > k \end{cases}
\end{aligned} \tag{3.10}$$

where $a = (1-\varepsilon)(2\Phi(k)-1)$ and k is defined by (3.9).

There are two modifications which can be made to ψ_1. First, since terms of order ρ^2 are being ignored, it seems reasonable to simplify $\psi_1(x)$ by eliminating the higher order terms (involved in the second term for $|x| > k$). Also note that for x large (positive) $\psi_1(x)$ decreases linearly to $-\infty$. Actually if $EX^2 < +\infty$, the probability that $\psi_1(x)$ is negative is less than the probability that $|X| \geqq c/\rho$, which is less than $c^{-2}\rho^2 EX^2$. Since we are only interested in first order terms it seems not unreasonable to truncate ψ_1 when it crosses the x-axis (in particular, this will yield an estimator of the form suggested by Hampel in [1]). Therefore, we will generally take the following function $\hat{\psi}_1$ as the minimax ψ-function:

$$\begin{aligned}
\hat{\psi}_1(x) &= x & |x| &\leqq k \\
&= k \operatorname{sgn} x - 2\rho a(x - k \operatorname{sgn} x) & k < |x| &< k' \\
&= 0 & |x| &\geqq k'
\end{aligned} \tag{3.11}$$

where k satisfies (3.9), $a = (1-\varepsilon)(2\Phi(k)-1)$ and

$$k' = k\,\frac{(1+2\rho a)}{2\rho a}\,. \tag{3.12}$$

We now give a rigorous theorem showing that $\hat{\psi}_1(x)$ is optimal in a certain sense:

THEOREM 3.1. *Let ψ of the form* (3.11) *be fixed. For each vector* $\mathbf{C} = (c_1, c_2, c_3, c_4)$ *let* $\mathscr{F}(\mathbf{c})$ *be the set of distributions* G, *in* $\mathscr{F}_1$ *such that* $E\psi'(Y) \geqq c_4 > 0$, $EY^2 \leqq c_1$, *G has a density, g, which is twice continuously differentiable except on a closed set D of Lebesgue measure zero (with $0 \notin D$),* $\sup_x |g(x)| \leqq c_2$, *and for each u*

$$\int_{\mathrm{compl}(D(u))} |g''(x)|\, dx \leqq c_3$$

where for each u, $D(u)$ is a neighborhood of D with Lebesgue measure u. For $G \in \mathscr{F}(\mathbf{c})$ let $\{\hat{\theta}_n\}$ be any sequence of invariant estimators such that $n^{\frac{1}{2}}(\hat{\theta}_n - \theta) \rightarrow_D \mathscr{N}(0, \sigma^2(G))$ *and (using Theorem A.4) let $\sigma_0^2(G)$ denote the asymptotic variance of the M estimator defined by ψ. Then for sufficiently large* $\mathbf{c}$ *there is* $B = B(\mathbf{c}) < +\infty$ *and* $\rho_0 = \rho_0(\mathbf{c}) > 0$ *such that*

$$\sup_{G \in \mathscr{F}(\mathbf{c})} \sigma^2(G) \geqq \sup_{G \in \mathscr{F}(\mathbf{c})} \sigma_0^2(G) - B\rho^2 \qquad \textit{for} \quad |\rho| \leqq \rho_0 .$$

PROOF. We first extend the derivation of (2.11) to distributions in $\mathscr{F}(\mathbf{c})$. For each sufficiently small ρ, let g_ρ be a density such that $g_\rho(x) = g(x)$ if $x \notin D(\rho)$,

$$|g(x) - g_\rho(x)| \leqq \rho \quad \text{for} \quad x \in D(\rho) \qquad \text{and} \qquad \int |g_\rho''(x)|\, dx \leqq b$$

where b depends only on $\mathbf{c}$ (this is possible since $|g_\rho''(x)| \leqq b/\rho$ on a set of measure ρ). Let $Y^{(\rho)}$ denote a random variable with density g_ρ and note that for any bounded function f,

$$(3.13) \qquad |Ef(Y) - Ef(Y^{(\rho)})| \leqq \int_{D(\rho)} |f(x)|\,|g(x) - g_\rho(x)|\, dx \leqq b_1\rho^2 .$$

From the appendix,

$$\sigma_0^2(G) = \frac{E\psi^2(X_1) + 2E\psi(X_1)\psi(X_2) + 2E\psi(X_1)\psi(X_3)}{(E\psi'(X_1))^2} .$$

Now let b denote a generic constant depending on $\mathbf{c}$ (and perhaps also on the bounds for ψ and ψ'). Then, from (3.13)

$$\begin{aligned} E\psi^2(X_1) = E\psi^2(Y_1 + \rho(Y_0 + Y_2)) &= E\psi^2(Y_1^{(\rho)} + \rho(Y_0^{(\rho)} + Y_2^{(\rho)})) + b\mathscr{O}(\rho^2) \\ &= E\psi^2(Y_1^{(\rho)}) + b\mathscr{O}(\rho^2) \\ &= E\psi^2(Y_1) + b\mathscr{O}(\rho^2) \end{aligned}$$

where the third equality uses the proof of Theorem 2.1. Similarly,

$$\begin{aligned} E\psi(X_1)\psi(X_2) &= -2\rho EY_1\psi(Y_1)E\psi'(Y_1) + b\mathscr{O}(\rho^2) \\ E\psi(X_1)\psi(X_3) &= b\mathscr{O}(\rho^2) \\ E\psi'(X_1) &= E\psi'(Y_1) + b\mathscr{O}(\rho^2) . \end{aligned}$$

Thus, $\sigma_0^2(G)$ is given by (2.11) with error uniform on $\mathscr{F}(\mathbf{c})$. In Section 4 it is shown that $\sigma_0^2(G)$ is maximized over $\mathscr{F}_1$ by a distribution, F^*, concentrating all the contamination at a point mass on the first break point of ψ. A direct calculation shows that the contribution to $\sigma_0^2(F^*)$ from integrals over $|Y| \leqq k/\rho$ equals the contribution to $\sigma_0^2(F_0)$ where F_0 is the least favorable distribution for $\mathscr{F}_1$ given by Huber. Since both the normal distribution and F_0 have finite variances, $P\{|Y| \geqq k/\rho\} \leqq b\rho^2$. Therefore, since $F_0 \in \mathscr{F}(\mathbf{c})$ for $\mathbf{c}$ sufficiently large, there is $B_1 = B(\mathbf{c})$ and $\rho_0 = \rho_0(\mathbf{c})$ with

$$\begin{aligned}\sup_{G \in \mathscr{S}(\mathbf{c})} \sigma_0^2(G) &= \sigma_0^2(F_0) + b\mathscr{O}(\rho^2) \\ &\leqq (1 + 4\rho)\,\frac{E\psi^2(U)}{(E\psi'(U))^2} + B_1\rho^2 \qquad \text{for} \quad |\rho| \leqq \rho_0\,,\end{aligned}$$

where U has distribution F_0 and the last equality uses (3.8).

Now by Theorem 1.1 (since $d = 1 + 2\rho$),

$$\sup_{G \in \mathscr{S}(\mathbf{c})} \sigma^2(G) \geqq (1 + 2\rho)^2 \frac{E\psi^2(U)}{E\psi'(U)^2}$$

and the theorem follows.

We conclude this section by noting that a linear combination of order statistics can be found which is equivalent to ψ_1 (3.10) and is also asymptotically minimax to first order terms. Since the influence curve for ψ_1 is proportional to ψ_1, the same shaped influence curve can be obtained by choosing a J-function (defining the linear combination of order statistics as in [5]) as follows:

$$(3.14) \qquad \begin{aligned} J(t) &= -2\rho & & t < \gamma \quad \text{or} \quad t \geqq 1 - \gamma \\ &= \frac{1 + 4\gamma\rho}{1 - 2\gamma} & & \gamma < t < 1 - \gamma \end{aligned}$$

where γ is an appropriate constant, $0 < \gamma < 1$. The influence curve for this estimator can be found to be

$$(3.15) \qquad \begin{aligned} I_F(x) &= \frac{1 + 4\rho\gamma}{1 - 2\gamma}\, x & & |x| < k(\gamma) \\ &= \frac{1 + 2\rho}{1 - 2\gamma}\, k \operatorname{sgn} x - 2\rho x & & |x| \geqq k(\gamma) \end{aligned}$$

where $k(\gamma) = -F^{-1}(\gamma)$. In the next section, we show that the approximate asymptotic variance (2.7) is maximized over $\mathscr{F}_1$ by choosing F^* to be $\mathscr{N}(0, 1)$ with probability $(1 - \varepsilon)$ and a 2-point distribution concentrated at $\pm k(\gamma)$ with probability ε. Thus, the maximum over $\mathscr{F}_1$ is achieved for $k(\gamma)$ satisfying

$$1 - 2\gamma = (1 - \varepsilon)(2\Phi(k(\gamma)) - 1)\,.$$

It follows that if γ is chosen so that $k(\gamma) = k$ for ψ_1 (given by (3.10)) then a direct calculation shows that the influence curve (3.15) at F^* is exactly the same as that for ψ_1 at its least favorable distribution. Therefore, the linear combination of order statistics defined by the J-function (3.14) is also minimax over $\mathscr{F}_1$ for the first order asymptotic variance (2.7).

4. Comparing estimators by maximum asymptotic variances over $\mathscr{F}_1$. This section compares the estimators derived in the previous section with several estimators suggested in Andrews et al. [1]. The comparison will be in terms of the approximate asymptotic variance (2.7), maximized over the class $\mathscr{F}_1$ of ε-contaminates of a normal distribution. We first show how to find the maximum variance for estimators of the last section. We then discuss the maximization for more general estimators, and conclude with a discussion of numerical results.

In particular, first consider M-estimators (including $\hat{\psi}_1$ given in (3.11)) of the form suggested by Hampel in [1]: ψ is piecewise linear (and odd) with break-points at $\pm a$, $\pm b$ and $\pm c$ and with slope one about the origin, slope zero between a and b, slope negative between b and c, and equal to zero beyond c. For M-estimators in ε-contaminated situations, the relevent expression to maximize is

$$(4.1) \qquad \sigma^2 = \frac{(1-\varepsilon)E\psi^2(Z) + \varepsilon E\psi^2(X)}{[(1-\varepsilon)E\psi'(Z) + \varepsilon E\psi'(X)]^2} + 4\rho \frac{(1-\varepsilon)EZ\psi(Z) + \varepsilon EX\psi(X)}{(1-\varepsilon)E\psi'(Z) + \varepsilon E\psi'(X)}$$

where $Z \sim \mathscr{N}(0, 1)$ and we want to maximize (4.1) over all distributions for X. For Hampel estimators, $\psi'(x)$ is either zero, one or negative; and, hence, both denominators in (4.1) are minimized by any distribution concentrated on $[b, c] \cup [-c, -b]$. Since $\psi(x) = 0$ for $|x| \geqq c$, the numerators are also maximized by such distributions. Furthermore, since such distributions make the denominators constant (say at a value, D), it remains to choose a distribution for X which maximizes

$$(4.2) \qquad \varepsilon E\psi^2(X)/D^2 + 4\rho\varepsilon EX\psi(X)/D = \frac{\varepsilon}{D^2} E\{\psi^2(X) + 4\rho DX\psi(X)\}\,.$$

The quantity (4.2) is maximized by simply choosing X to be concentrated at a value maximizing $\psi^2(x) + 4\rho Dx\psi(x)$ that is at some value $\pm x$ for $b \leqq x \leqq c$. Since ψ is piecewise linear, the function is maximized either at $x = b$ or at a unique point in (b, c) which can be easily found. In fact for ρ small it is easy to see that $(d/dx)(\psi^2(x) + 4\rho Dx\psi(x))$ is negative on (b, c) and, hence, the maximum is at $x = b$.

The maximization is somewhat harder for the linear combinations of order statistics like (3.14), since the break-points in the influence curve depend on the distribution. Actually, as (3.14) is written, the maximum variance is infinity since the influence curve (3.15) is not bounded. However, if distributions are restricted to those not concentrating too much probability on the set where $|x|$ is large, (3.14) will be approximately equivalent to an estimator which trims a very small proportion of extreme order statistics (instead of giving them negative weight); and the maximum variance over a moderately large interval should be a reasonable value for comparing estimators.

We now want to maximize (2.7) with I_F given by (3.15), and F an ε-contaminate. Here, γ is a constant defining the estimator, and $I_F(x)$ is piecewise linear with positive slope for $|x| < k(\gamma)$ and negative slope for $x \geqq k(\gamma)$ (where $k(\gamma) = -F^{-1}(\gamma)$). The asymptotic variance here is

$$(4.3) \qquad \begin{aligned} \sigma^2 = {} & (1-\varepsilon)EI_F{}^2(Z) + \varepsilon EI_F{}^2(X) \\ & + 4\rho[(1-\varepsilon)EI_F{}'(Z) + \varepsilon EI_F{}'(X)][(1-\varepsilon)EZI_F(Z) + \varepsilon EXI_F(X)]\,, \end{aligned}$$

where $Z \sim \mathscr{N}(0, 1)$ and σ^2 is to be maximized over the distribution of X. Assuming F is continuous and letting $Y \sim F$, then the probability is exactly 2γ that $|Y| \geqq k$. Therefore

$$EI_F'(Y) = 2\gamma(-2\rho) + (1 - 2\gamma)\left(\frac{1 + 4\rho\gamma}{1 - 2\gamma}\right) = 1 \tag{4.4}$$

independent of what F is. Thus, we must maximize

$$EI_F^2(X) + 4\rho EXI_F(X) . \tag{4.5}$$

Clearly, to maximize (4.5), F should be chosen so that $k(\gamma)$ (and, hence, I_F itself) is as large as possible. If $\gamma \leqq \varepsilon/2$ then $k(\gamma)$ may be chosen to be as large as possible and, hence, σ^2 will approach ∞ as F puts more probability in its tails. If $\gamma > \varepsilon/2$, the maximum value for $k(\gamma)$ is $k_0 = -\Phi^{-1}((1/(1 - \varepsilon))(\gamma - \varepsilon/2))$ (where Φ is the $\mathscr{N}(0, 1)$ cdf). So here σ^2 is maximized by concentrating probability at $\pm x_0$ where x_0 maximizes $I_F^2(x) + 4\rho x I_F(x)$. This function is proportional to x^2 for $|x| \leqq k_0$ and is a different quadratic for $x > k_0$. Either the tail quadratic has a unique maximum which will define the maximum variance, or the function has a local maximum at k_0 which, although not a global maximum, will in this case be taken to define the maximum variance. In the tabulated values later presented, this local maximum will represent a maximum at least over the range $-20 \leqq x \leqq 20$.

For more general M-estimators, the method of moment spaces may be used to show that if $\rho = 0$, then the maximum asymptotic variance over $\mathscr{F}_1$ is achieved at a distribution contaminated by a two-point distribution (concentrating probability $\frac{1}{2}$ at $\pm x$). In particular, let $\mathscr{F}$ be the family of all symmetric distributions, and let ψ be an odd function defining an M-estimator. Consider the affine transformation $A: \mathscr{F} \to R^2$ defined to have coordinates

$$A_1(F) = E_G\psi'(X) , \qquad A_2(F) = E_G\psi^2(X) , \tag{4.6}$$

where G is a function of F: $G = (1 - \varepsilon)\Phi + \varepsilon F$. Since $\mathscr{F}$ is a compact convex set for which the extreme points are the two-point distribution, the set $B = A(\mathscr{F})$ is a compact convex subset of R^2 for which the extreme points are images of two-point distributions. In fact the upper boundary of B (under general smoothness conditions on ψ) will just be the image of the curve $\{F_x : 0 \leqq x \leqq a\}$, where F_x is the two-point distribution concentrating probability $\frac{1}{2}$ at $\pm x$. Figure 1 gives an example of B for $\psi(x) = \sin \frac{1}{2}x$ (for $|x| \leqq 2\pi$, and $\psi(x) = 0$ for $|x| > 2\pi$) where the upper boundary corresponds to $\{F_x : 0 \leqq x \leqq 2\pi\}$. To maximize the asymptotic variance, we must find the point (x, y) in B which maximizes y/x^2. Considering the convex functions $y = bx^2$ as b increases, we see that the maximum value occurs at the value of b^* for which the convex curve $y = b^*x^2$ is tangent to the concave upper boundary of B. Since this point of tangency is a unique point corresponding to a two-point distribution, the maximization problem is reduced to finding the maximum of a real valued function of a single real value.

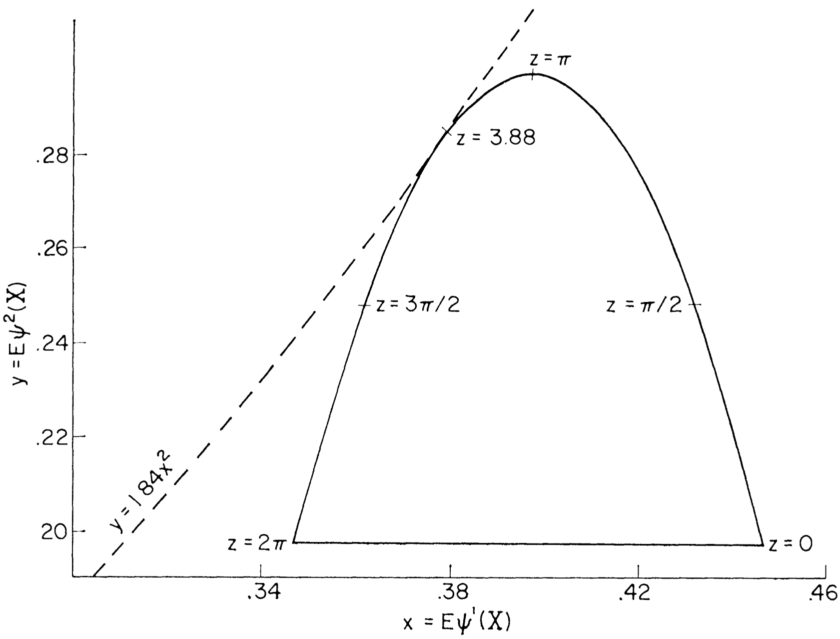

FIG. 1. Boundary of moment space for $\Psi(w) = \sin \frac{1}{2}w$ for $|w| \leqq 2\pi$ and $\Psi(w) = 0$ for $|w| > 2\pi$ under $(1 - \varepsilon)\mathcal{N}(0, 1) + \varepsilon\delta(\pm z)$.

To maximize the asymptotic variance (2.11) for $\rho \neq 0$, a third coordinate $A_3(F) = E_G X\psi(X)$ (again with $G = (1 - \varepsilon)\Phi + \varepsilon F$) must be considered. Again if ψ is smooth enough, the set B will be the convex hull of the curve corresponding to $\{F_x : 0 \leqq x \leqq a\}$. Once again, it is likely that the maximization over $\mathcal{F}_1$ is achieved at a distribution contaminated by a two-point distribution. Thus, again, maximization can be carried out over a single real parameter; that is, we need only consider maximizing the following function over $x > 0$:

$$V = [(1 - \varepsilon)E\psi^2(X) + \varepsilon\psi^2(x)]/D^2(x) + 4\rho[(1 + \varepsilon)EX\psi(X) + \varepsilon x\psi(x)]/D(x) \tag{4.7}$$

where $D(x) = (1 - \varepsilon)E\psi'(X) + \varepsilon\psi'(x)$ and where expectation is under $\mathcal{N}(0, 1)$.

Admittedly, this statement has not been proven. The best this argument provides is that the maximum must occur at a distribution contaminated by a convex combination of at most three two-point distributions. I believe that maximization of (4.7) over x suffices for all the cases considered here; in any event, it should yield an approximation good enough for the comparisons made here.

In the calculations listed here, only the case $\rho \geqq 0$ is considered. From (2.7) it is clear that $\rho < 0$ only reduces the variance. Thus, it is most important to have protection against the presence of positive correlation.

Values of maximum asymptotic variance are listed in Table 1 for a number of estimators as a function of the contaminating fraction: $\varepsilon = .05, .1, .2$, and $.3$, and the correlation $\rho = 0$, .1, .2 and .3. The optimal estimators (3.11) and (3.14) will be considered first. The M-estimator, (3.11), depends on an assumed value $\hat{\varepsilon}$ for the contaminating fraction and $\hat{\rho}$ for the correlation. Since they are Hampel-type estimators, they will be denoted by $H(\hat{\varepsilon} = \cdot, \hat{\rho} = \cdot)$ for $\hat{\varepsilon} = .05$, .1, and .2 and $\hat{\rho} = .05$ and .1. One interesting conclusion should be noted at this point. Calculations were also carried out for $\hat{\rho} = .25$ and they show that, over the range of ε and ρ in Table 1, $\hat{\rho} = .1$ is always a uniformly better assumption than $\hat{\rho} = .25$ (in terms of asymptotic variance).

The optimal linear combinations of order statistics, (3.14), depend on assumed values $\hat{\gamma}$ for the trimming proportion and $\hat{\rho}$ for the correlation. They are denoted by $L(\hat{\gamma} = \cdot, \hat{\rho} = \cdot)$ and listed in Table 1 for $\hat{\gamma} = .1$, .2, .25 and .4 and $\hat{\rho} = .05$, and .1. Again, the choice $\hat{\rho} = .1$ is always preferable to the choice $\hat{\rho} = .2$ over the range of ε and ρ in Table 1.

These optimal estimators may be compared with other estimators discussed in [1]. In particular, calculations were carried out for the Hampel estimators $H(a, b, c)$ which are M-estimators where the ψ-function is piecewise linear with breakpoints at $\pm a$, $\pm b$, and $\pm c$. Over the range in Table 1, $H(1.2, 3.5, 8)$ is

TABLE 1

Maximum asymptotic variance for contaminating fraction ε and correlation ρ.

Part (1). *Optimal M-estimates for assumed $\hat{\varepsilon}$ and $\hat{\rho}$*

estimator	ε	$\rho = 0$	$\rho = .1$	$\rho = .2$	$\rho = .3$
$H(\hat{\varepsilon} = .05, \hat{\rho} = .05)$	.05	1.28	1.77	2.35	2.93
	.1	1.56	2.14	2.91	3.70
	.2	2.33	3.08	4.31	5.59
	.3	3.56	4.48	6.26	8.18
$H(\hat{\varepsilon} = .05, \hat{\rho} = .1)$	.05	1.31	1.77	2.25	2.75
	.1	1.62	2.14	2.70	3.32
	.2	2.50	3.17	3.89	4.78
	.3	3.98	4.87	5.77	6.95
$H(\hat{\varepsilon} = .1, \hat{\rho} = .05)$	.05	1.30	1.78	2.33	2.89
	.1	1.55	2.11	2.83	3.57
	.2	2.24	2.94	4.07	5.24
	.3	3.32	4.17	5.79	7.52
$H(\hat{\varepsilon} = .1, \hat{\rho} = .1)$	.05	1.34	1.78	2.25	2.74
	.1	1.62	2.11	2.65	3.24
	.2	2.42	3.03	3.70	4.53
	.3	3.74	4.52	5.33	6.43
$H(\hat{\varepsilon} = .2, \hat{\rho} = .05)$	.05	1.35	1.83	2.37	2.92
	.1	1.58	2.13	2.84	3.56
	.2	2.21	2.90	3.99	5.11
	.3	3.18	4.01	5.58	7.22
$H(\hat{\varepsilon} = .2, \hat{\rho} = .1)$	.05	1.40	1.84	2.30	2.79
	.1	1.67	2.13	2.67	3.24
	.2	2.40	2.95	3.62	4.41
	.3	3.59	4.26	5.04	6.11

Part (2). *Optimal linear combination of order statistics for assumed* $\hat{\gamma}$ *and* $\hat{\rho}$

estimator	ε	$\rho = 0$	$\rho = .1$	$\rho = .2$	$\rho = .3$
$L(\hat{\gamma} = .1, \hat{\rho} = .05)$	.05	1.28	1.77	2.36	2.94
	.1	1.59	2.22	3.08	3.96
	$\geqq .2$	∞	∞	∞	∞
$L(\hat{\gamma} = .1, \hat{\rho} = .1)$	.05	1.31	1.77	2.25	2.76
	.1	1.63	2.17	2.78	3.46
	$\geqq .2$	∞	∞	∞	∞
$L(\hat{\gamma} = .2, \hat{\rho} = .05)$	.05	1.35	1.81	2.34	2.88
	.1	1.57	2.11	2.81	3.53
	.2	2.22	3.01	4.27	5.56
	.3	3.52	4.88	7.36	9.91
$L(\hat{\gamma} = .2, \hat{\rho} = .1)$	.05	1.40	1.83	2.29	2.77
	.1	1.64	2.12	2.65	3.24
	.2	2.35	2.96	3.74	4.65
	.3	3.77	4.66	5.97	7.60
$L(\hat{\gamma} = .25, \hat{\rho} = .05)$	.05	1.41	1.86	2.38	2.91
	.1	1.62	2.14	2.81	3.50
	.2	2.20	2.93	4.06	5.22
	.3	3.19	4.30	6.28	8.33
$L(\hat{\gamma} = .25, \hat{\rho} = .1)$	.05	1.47	1.90	2.35	2.83
	.1	1.71	2.16	2.69	3.25
	.2	2.35	2.91	3.63	4.47
	.3	3.45	4.18	5.27	6.62
$L(\hat{\gamma} = .4, \hat{\rho} = .05)$	.05	1.65	2.09	2.60	3.11
	.1	1.87	2.36	2.99	3.63
	.2	2.42	3.06	4.05	5.07
	.3	3.25	4.11	5.68	7.30
$L(\hat{\gamma} = .4, \hat{\rho} = .1)$	.05	1.77	2.18	2.63	3.10
	.1	2.02	2.45	2.95	3.49
	.2	2.67	3.13	3.79	4.55
	.3	3.63	4.15	5.05	6.15

Part (3). *Other M-estimators*

estimator	ε	$\rho = 0$	$\rho = .1$	$\rho = .2$	$\rho = .3$
$H(1.2, 3.5, 8)$	.05	1.31	1.83	2.36	2.88
	.1	1.61	2.28	2.95	3.62
	.2	2.52	3.55	4.59	5.62
	.3	4.16	5.73	7.30	8.87
$H(2.1, 4, 8.2)$	.05	1.40	2.00	2.60	3.20
	.1	1.92	2.76	3.59	4.43
	.2	3.62	5.08	6.54	8.00
	.3	7.18	9.62	12.05	14.48
sin, (4.8) $a = 1/2.1$	.05	1.39	1.99	2.59	3.19
	.1	1.88	2.72	3.56	4.41
	.2	3.46	4.97	6.49	8.02
	.3	7.19	9.93	12.67	15.41

cont.

Part (3) Continued

estimator	ε	$\rho = 0$	$\rho = .1$	$\rho = .2$	$\rho = .3$
sin, (4.8) $a = .5$	.05	1.37	1.95	2.54	3.13
	.1	1.84	2.64	3.46	4.27
	.2	3.35	4.78	6.23	7.67
	.3	7.00	9.58	12.16	14.75
sin, (4.8) $a = .6$	.05	1.33	1.87	2.41	2.96
	.1	1.73	2.44	3.16	3.88
	.2	3.08	4.29	5.50	6.71
	.3	6.73	8.85	10.98	13.11
sin, (4.8) $a = .7$	.05	1.32	1.84	2.36	2.87
	.1	1.70	2.36	3.02	3.68
	.2	3.04	4.12	5.19	6.27
	.3	7.32	9.11	10.97	12.85
sin, (4.8) $a = .8$	.05	1.35	1.85	2.35	2.85
	.1	1.73	2.35	2.98	3.61
	.2	3.18	4.18	5.18	6.18
	.3	9.09	10.49	12.15	13.87
Olshen, (4.9) $a = .5$	.05	1.84	2.28	2.73	3.17
	.1	2.22	2.72	3.22	3.72
	.2	3.40	4.05	4.69	5.34
	.3	5.77	6.65	7.52	8.40
Olshen, (4.9) $a = 2.0$	.05	1.37	1.83	2.30	2.77
	.1	1.61	2.15	2.69	3.24
	.2	2.31	3.03	3.77	4.51
	.3	3.48	4.46	5.47	6.49
Olshen, (4.9) $a = 3.0$	.05	1.31	1.79	2.27	2.76
	.1	1.56	2.12	2.70	3.29
	.2	2.25	3.03	3.85	4.68
	.3	3.39	4.49	5.63	6.79
Olshen, (4.9) $a = 4.0$	.05	1.29	1.78	2.28	2.79
	.1	1.54	2.14	2.75	3.38
	.2	2.26	3.11	4.00	4.92
	.3	3.42	4.64	5.91	7.22
exp, (4.10) $a = .10$	.05	1.30	1.81	2.32	2.84
	.1	1.63	2.27	2.92	3.57
	.2	2.64	3.64	4.64	5.65
	.3	4.65	6.22	7.79	9.37
exp, (4.10) $a = .125$	.05	1.31	1.81	2.31	2.81
	.1	1.63	2.24	2.86	3.48
	.2	2.62	3.55	4.49	5.42
	.3	4.64	6.09	7.54	8.99
exp, (4.10) $a = .15$	.05	1.33	1.82	2.31	2.80
	.1	1.65	2.24	2.84	3.44
	.2	2.65	3.55	4.43	5.32
	.3	4.75	6.12	7.49	8.87

uniformly better than any other Hampel estimator presented in [1] (including ADA, which is asymptotically equivalent to $H(1, 4.5, 8)$ for the distribution maximizing the asymptotic variance). Thus, $H(1.2, 3.5, 8)$ is listed; and, for purposes of comparison, $H(2.1, 4, 8.2)$ is also listed.

Lastly, three other M-estimators with the following ψ-functions are listed (where "a" denotes a parameter defining the estimator):

$$\psi_1(x) = \sin ax \quad |x| < \frac{\pi}{a} \qquad (4.8)$$
$$= 0 \quad |x| \geqq \frac{\pi}{a} ,$$

$$\psi_2(x) = \frac{x}{a + x^2} , \qquad (4.9)$$

$$\psi_3(x) = x \exp\{-ax^2\} . \qquad (4.10)$$

The estimator defined by ψ_1 (denoted by "sin, (4.8)") was suggested by Andrews [1] when $a = 1/2.1$. Asymptotic variances are listed for $a = 1/2.1$, .5, .6, .7, and .8. Numerous other values were used in calculations, but either $a = .6$, .7, or .8 was always uniformly better over the range in Table 1. The ψ_2 estimator (denoted by "Olshen, (4.9)") was suggested by Olshen when $a = .5$. Results are listed here for $a = .5$, 2, 3 and 4. The ψ_3 estimator is included since it seems a priori to be a reasonable alternative to ψ_1 and ψ_2. Results are listed for $a = .1$, .125, and .15, which appear to be optimal over the range in Table 1.

The main conclusions which can be drawn from Table 1 are the following:

(1) Relatively small amounts of correlation can drastically inflate asymptotic variances; and many estimators which appear to be quite good when $\rho = 0$ can be extremely poor when $\rho = .2$ or $\rho = .3$ (for example, consider $H(2.1, 4, 8.2)$, "sin" for $a = 1/2.1$, or "Olshen" for $a = .5$).

(2) The optimal M-estimators appear to be slightly better than the linear combinations of order statistics, and substantially better than any of the Hampel-type estimators suggested in [1] or any of the "sin" type or "exp" type estimators considered here. They are also substantially better than Olshen's suggestion with $a = .5$. In particular, if the contaminating fraction ε is thought to be moderate but not too large, the estimator $H(1, 1, 10)$ would probably be about as good as possible; whereas if ε is thought to be small, $H(1.5, 1.5, 10.0)$ should work well.

(3) The Olshen type estimators for a between 2.0 and 4.0 also seem to have extremely good asymptotic variance over the range of Table 1; and one such estimator might be preferable for actual use since it would be smooth. In particular, if ε is likely to be moderate, $a = 2.0$ is a good choice, and if ε is small, $a = 3.5$ is a good choice. These estimators are good probably because values of a which keep the asymptotic variance moderate in the independent case, also keep the slope from becoming too negative.

(4) The optimal linear combinations of order statistics also seem to be somewhat better than most of the other estimators considered in [1]. Since they are easy to calculate, they could also be strongly recommended in cases where small positive correlation is not entirely unexpected.

5. Third order expansion for asymptotic variance. The expansions of Theorem 2.1 can be carried out to higher orders in ρ in a straightforward though tedious manner. The only major complication is that a strengthening of the last part of (2.6) is required. To obtain an expansion to terms in ρ^3, the following assumption on the estimator, T, and the cdf G in model (1.1) will be made:

$$I_F(x) = I_G(x) + \rho^2 J_G(x) + \mathcal{O}(\rho^4) \tag{5.1}$$

uniformly in x where J_G depends on G but not on ρ. Equation (5.1) can be expected to hold whenever G and T are smooth. In particular, consider an M-estimator such that both Ψ and Ψ' are bounded. Under the regularity conditions on G in Theorem 5.1 below, the density f of X_i can be expanded as in (2.9) and (2.10) to obtain

$$f_X(x) = g(x) + \rho^2(EY^2)g''(x) + g^{(4)}(x)\mathcal{O}(\rho^4)\,.$$

Therefore,

$$E\Psi'(X) = E\Psi'(Y) + \rho^2(EY^2) \int \Psi'(x)g''(x)\,dx + \mathcal{O}(\rho^4)\,.$$

Thus,

$$\begin{aligned} I_F(x) &= \frac{\Psi(x)}{(E\Psi'(X))} = \frac{\Psi(x)}{(E\Psi'(Y))}\left(1 - \rho^2 \frac{EY^2 \int \Psi'(x)g''(x)\,dx}{E\Psi'(Y)} + \mathcal{O}(\rho^4)\right) \\ &= I_G(x) - c\rho^2\Psi(x) + \mathcal{O}(\rho^4) \end{aligned} \tag{5.2}$$

(for appropriate c); and, hence, (5.1) holds.

THEOREM 5.1. *Under model* (1.1), *assume that F (the marginal of X_i) and the estimator T satisfy* (5.1) *and the condition* $|I_F(x)| \leq C_1$. *Assume also that G has finite fourth moments and a density g with absolutely integrable fourth derivative* (*equivalently*, $\int u^4|\varphi_G(u)|\,du < +\infty$). *Let* $c = EY^2$. *Then*

$$\begin{aligned} \sigma^2 = {} & EI_G^2(Y) - 4\rho EYI_G(Y) \int I_G(x)g'(x)\,dx \\ & + \rho^2\{2EI_G(Y)J_G(Y) + c \int I_G^2(x)g''(x)\,dx + 2c[\int I_G(x)g'(x)\,dx]^2\} \\ & - 2\rho^3\{2[EYI_G(Y) \int J_G(x)g'(x)\,dx + EYJ_G(Y) \int I_G(x)g'(x)\,dx] \\ & + \int I_G(x)[2g'(x) + xg''(x)]\,dx \int I_G(x)[2xg(x) + (x^2 + c)g'(x)]\,dx \\ & + (\tfrac{1}{3}EY^3I_G(Y) + cEYI_G(Y)) \int I_G(x)g'''(x)\,dx\} + \mathcal{O}(\rho^4)\,. \end{aligned} \tag{5.3}$$

The proof is a straightforward generalization of the proof of Theorem 2.1 and is not given here. Again integration by parts will yield an equivalent expansion which will hold whenever I_G and J_G are smooth enough. Furthermore, inserting (5.2) in (5.3) and integrating by parts in the terms involving derivatives of g

TABLE 2
Coefficients of ρ^n *in approximation* $\sigma^2 = V_1 + V_2\rho + V_3\rho^2 + V_4\rho^3$
under $(1-\varepsilon)\mathcal{N}(0, 1) + \varepsilon\mathcal{N}(0, w^2)$

estimator	ε	w	V_1	V_2	V_3	V_4
$H(\hat{\varepsilon} = .1, \rho = .05)$	.1	3	1.31	5.17	4.87	4.15
		5	1.35	5.80	9.54	7.29
	.3	3	1.93	8.09	11.53	12.98
		5	2.13	10.59	31.85	45.70
$H(1.0, 4.5, 8.0)$	.1	3	1.33	5.13	4.89	4.79
		5	1.36	5.30	9.60	11.23
	.3	3	1.98	7.97	11.78	14.96
		5	2.17	8.84	33.02	52.45
$H(2, 2, 10)$	.1	3	1.28	5.33	4.13	2.02
		5	1.31	5.67	8.14	4.52
	.3	3	2.03	8.60	9.18	5.97
		5	2.26	10.12	25.81	22.29

gives the following expansion for the case of M-estimators:

$$\sigma^2 = \frac{E\psi^2(Y)}{(E\psi'(Y))^2} + 4\rho\frac{EY\psi(Y)}{E\psi'(Y)}$$
$$+ 2\rho^2 c\left\{-\frac{E\psi'''(Y)E\psi^2(Y)}{(E\psi'(Y))^3} + \frac{E\psi''(Y)\psi(Y) + E\psi'(Y)^2}{(E\psi'(Y))^2} + 1\right\}$$
$$+ \frac{2\rho^3}{(E\psi'(Y))^2}\{E\psi'''(Y)E(\tfrac{1}{3}Y^3 - 3cY)\psi(Y) + EY\psi''(Y)E(Y^2 + c)\psi'(Y)\}$$

where, as before, Y has cdf G and $c = EY^2$.

Thus, we have an approximation of the form

$$\sigma^2 = V_1 + V_2\rho + V_3\rho^2 + V_4\rho^3 .$$

Table 2 lists values for V_1, V_2, V_3 and V_4 for a small number of Hampel-type estimators for distributions of the form $(1-\varepsilon)\mathcal{N}(0, 1) + \varepsilon\mathcal{N}(0, w^2)$, where $w = 3$ and 5 and $\varepsilon = .1$ and .3. The estimators presented represent the worst cases among the Hampel type estimators of Section 4 in the sense that V_3 and V_4 tend to be largest for these estimators. Results for smooth estimators in situations in which the contamination is degenerate should be quite similar. However, relative errors should be somewhat smaller when both the estimator and the distribution are sufficiently smooth. The basic conclusions to be drawn from Table 2 are the following:

(1) The first term approximation is an underestimate, and in fact, σ^2 may be substantially larger when both ρ and ε are much larger than .1. This would be expected from the discussion in Section 1.

(2) For $\rho \leqq .1$, the first term approximation is probably quite adequate, having a maximum relative error in the tabulated range of about 10%. The approximation is probably adequate for rough comparisons as long as ρ or ε is less than .2.

(3) Since values for V_3 and V_4 are of similar magnitude, the comparisons of Section 4 are legitimate. However, actually variances listed in Table 1 may be somewhat in error for $\rho = .3$ and $\varepsilon = .3$.

APPENDIX

Proofs of consistency and asymptotic normality:

LEMMA A.1. *If $X_1, X_2, \cdots$ is a sequence of bounded m-dependent random variables then $\bar{X}_n - E\bar{X}_n \rightarrow_P 0$.*

Throughout the rest of the appendix we consider observations $X_i = Z_i + \theta$ ($i = 1, 2, \cdots$) where $\{Z_1, Z_2, \cdots\}$ is an m-dependent sequence of random variables, and all probability statements are under a fixed value θ_0.

THEOREM A.2. *Assume Ψ is continuous, bounded and $E_\theta \Psi(X)$ is strictly increasing in a neighborhood of $\theta = 0$ with $E_0 \Psi(X) = 0$. Let $\tilde{\theta}_n \rightarrow_P \theta_0$, and let $\hat{\theta}_n$ be the root of the equation $0 = \sum_{i=1}^n \Psi(X_i - \theta)$ nearest to $\tilde{\theta}_n$ (and larger if there are two roots equally distant from $\tilde{\theta}_n$). For definiteness, let $\hat{\theta}_n = 0$ if there is no root. Then $\hat{\theta}_n \rightarrow_P \theta_0$.*

COROLLARY A.3. *Let $a > 0$ and $b > 0$ be fixed and let $d(\theta)$ be a strictly increasing function on $(-b, b)$ with $d(0) = 0$. Let $\mathscr{S}$ be the set of continuous functions Ψ satisfying*

(i) $|\Psi(x)| \leqq a$ *for all* x;

(ii) $E_\theta \Psi(X) > d(\theta)$ *for* $0 < \theta < b$ *and* $E_\theta \Psi(X) < d(\theta)$ *for* $-b < \theta < 0$.

Then $\hat{\theta}_n \rightarrow_P \theta_0$ uniformly for $\Psi \in \mathscr{S}$.

THEOREM A.4. *Assume the hypotheses of Theorem A.2; and suppose in addition that Ψ is differentiable except on a closed set D with zero Lebesgue measure, Ψ' is uniformly continuous and bounded off any neighborhood of D, $0 \notin D$, and $E_0 \Psi'(X) = c > 0$. Suppose furthermore that the marginal distribution of X_i under $\theta = 0$ has a bounded density in a neighborhood of D. Define*

$$\sigma^2 = \lim_{n\to\infty} \frac{1}{n} \operatorname{Var}\left(\textstyle\sum_{i=1}^n \Psi(X_i)\right).$$

Then if $\hat{\theta}_n$ is the root of $0 = \sum_{i=1}^n \Psi(X_i - \theta)$ defined in Theorem A.2, $n^{\frac{1}{2}}(\hat{\theta}_n - \theta) \rightarrow_D \mathscr{N}(0, \sigma^2/c^2)$.

PROOF. There exist functions $\Psi_k(x)$ ($k = 1, 2, \cdots$) which equal Ψ outside a small neighborhood $D(\delta/2)$ (where $D(\delta) \supset D$ is chosen below depending on the distribution), are bounded (uniformly in k), are continuously differentiable with $|\Psi_k'(x)| \leqq B$ for all x, and such that (i) $|\Psi_k(x) - \Psi(x)| \leqq 1/k$ for all x, (ii) $E_0 \Psi_k'(X) \rightarrow E_0 \Psi'(X)$, and (iii) $\sigma_k \rightarrow \sigma$ as $k \rightarrow \infty$ where σ_k^2 is defined analogously to σ^2 above. Now note that since $\Psi_k = \Psi$ in a neighborhood of zero, all Ψ_k lie in the same set, $\mathscr{S}$, of the form described in Corollary A.2. Hence, if $\hat{\theta}_n^k$ satisfy $0 = \sum_{i=1}^n \Psi_k(X_i - \theta)$ (and, say, are the nearest roots to $\hat{\theta}_n$) then $\hat{\theta}_n^k \rightarrow_P \theta_0$ uniformly in k.

Using the standard Taylor series expansion,

$$(\text{A.1}) \qquad n^{\frac{1}{2}}(\hat{\theta}_n{}^k - \theta_0) = -\frac{1}{n^{\frac{1}{2}}} \sum_{i=1}^n \Psi_k(X_i - \theta)\Big/\frac{1}{n} \sum_{i=1} \Psi_k'(X_i - \tilde{\theta}_n{}^k)$$

where $\tilde{\theta}_n{}^k \to_P \theta_0$. It is first shown that

$$(\text{A.2}) \qquad \frac{1}{n} \sum_{i=1}^n \Psi_k'(X_i - \tilde{\theta}_n{}^k) \to E_0 \Psi_k'(X) \qquad \text{uniformly in} \quad k\,.$$

Let $\varepsilon > 0$ be given. Choose N_0 and an open set $D(\delta) = \{x : \inf_{y \in D} |y - x| < \delta\}$ such that for $n \geqq N_0$,

$$p\left\{\left|\frac{M}{n}\right| \geqq \frac{\varepsilon}{6B}\right\} \leqq \frac{\varepsilon}{3}$$

where M is the number of $\{X_i - \theta\}$ lying in $D(\delta)$. This follows from Lemma A.1 since the distribution of X has a bounded density on $D(\delta) + \theta$ and, hence, the probability of a single observation in $D(\delta) + \theta$ is proportional to the Lebesgue measure of $D(\delta)$ (which can be made arbitrarily small by choosing δ small). Now since $\Psi_k(u) = \Psi(u)$ for $u \notin D(\delta)$ choose $\delta' < \delta/2$ such that if $|u - v| < \delta'$ and $u \notin D(\delta)$ then $|\Psi_k(u) - \Psi_k(v)| < \varepsilon/3$ for all k. Choose $N > N_0$ such that for $n \geqq N$, and for all k, $P\{|\tilde{\theta}_n{}^k - \theta_0| \geqq \delta\} \leqq \varepsilon/3$ and

$$P\left\{\frac{1}{n}\left|\sum_{i=1}^n \Psi_k'(X_i - \theta_0) - E_0 \Psi_k'(X)\right| \geqq \frac{\varepsilon}{3}\right\} \leqq \frac{\varepsilon}{3}$$

(this last inequality uses Lemma A.1 again). Then, letting $C = \{i : X_i - \theta_0 \notin D(\delta)\}$,

$$\begin{aligned} P&\left\{\left|\frac{1}{n} \sum_{i=1}^n \Psi_k'(X_i - \tilde{\theta}_n{}^k) - E_0 \Psi_k'(X)\right| \geqq \varepsilon\right\} \\ &\leqq P\left\{\frac{1}{n}\left|\sum_C \Psi_k'(X_i - \tilde{\theta}_n{}^k) - \sum_C \Psi_k'(X_i - \theta_0)\right| \geqq \frac{\varepsilon}{3}\right\} \\ &\quad + P\left\{\frac{1}{n} \sum_{C^c} (|\Psi_k'(X_i - \tilde{\theta}_n{}^k)| + |\Psi_k'(X_i - \theta_0)|) \geqq \frac{\varepsilon}{3}\right\} \\ &\quad + P\left\{\left|\frac{1}{n} \sum_{i=1}^n \Psi_k'(X_i - \theta_0) - E_0 \Psi'(X)\right| \geqq \frac{\varepsilon}{3}\right\} \\ &\leqq P\{|\tilde{\theta}_n{}^k - \theta_0| \geqq \delta\} + P\left\{\frac{2M \sup |\Psi_k'(x)|}{n} \geqq \frac{\varepsilon}{3}\right\} + \frac{\varepsilon}{3} \leqq \varepsilon\,. \end{aligned}$$

Hence, (A.2) holds.

Now by the Berry-Esseen theorem for m-dependent random variables (see Stein [6]),

$$\frac{1}{n^{\frac{1}{2}}} \sum_{i=1}^n \Psi_k(X_i - \theta_0) \to_D \mathcal{N}(0, \sigma_k{}^2) \qquad \text{uniformly in} \quad k\,.$$

Therefore, using equations A.1 and A.2, $n^{\frac{1}{2}}(\hat{\theta}_n{}^k - \theta_0) \to_D \mathcal{N}(0, \tilde{\sigma}_k{}^2)$ uniformly in k where

$$\tilde{\sigma}_k{}^2 = \sigma_k{}^2/(E\Psi_k'(X))^2\,.$$

To complete the proof of Theorem A.4, let $\varepsilon > 0$ be given. By choice of Ψ_k,

$$\frac{1}{k} \geqq \left|\frac{1}{n} \sum_{i=1}^n \Psi(X_i - \hat\theta_n) - \frac{1}{n} \sum_{i=1}^n \Psi_k(X_i - \hat\theta_n)\right| = \frac{1}{n} |\sum_{i=1}^n \Psi_k(X_i - \hat\theta_n)| ,$$

and again using a Taylor series expansion

$$\frac{1}{n} \sum_{i=1}^n \Psi_k(X_i - \theta_n) = (\hat\theta_n - \hat\theta_n{}^k) \frac{1}{n} \sum_{i=1}^n \Psi_k'(X_i - \tilde\theta_n{}^k)$$

or

$$|\hat\theta_n - \hat\theta_n{}^k| \leqq \left|\frac{1}{n} \sum_{i=1}^n \Psi_k(X_i - \hat\theta_n)\right| \Big/ \left|\frac{1}{n} \sum_{i=1}^n \Psi_k'(X_i - \tilde\theta_n{}^k)\right|$$

$$\leqq \frac{1}{k|(1/n) \sum_{i=1}^n \Psi_k'(X_i - \tilde\theta_n{}^k)|} .$$

Now, as above, there is N_1 such that for $n \geqq N_1$,

$$P\left\{\frac{1}{n} \sum_{i=1}^n \Psi_k'(X_i - \tilde\theta_n{}^k) \geqq \tfrac{2}{3} E_0 \Psi_k'(X)\right\} \geqq 1 - \frac{\varepsilon}{6} \qquad \text{for all} \quad k .$$

Thus, since $E_0 \Psi_k'(X) \to c$, there is K such that for $k \geqq K$

$$P\left\{|\hat\theta_n - \hat\theta_n{}^k| \leqq \frac{2}{ck}\right\} \geqq 1 - \frac{\varepsilon}{6} \qquad \text{for} \quad n \geqq N_1 .$$

Choose $N_2 > N_1$ so that $|P\{n^{\frac12}(\hat\theta_n{}^k - \theta_0) \leqq y\} - \Phi(y/\tilde\sigma_k)| \leqq \varepsilon/6$ for all y for $n \geqq N_2$; and (since $\tilde\sigma_k \to \sigma/c$ as $k \to \infty$) for each n choose $k = k(n) > K$ so that for all x

$$\left|\Phi\left(\frac{x + 2n^{\frac12}/ck}{\tilde\sigma_k}\right) - \Phi\left(\frac{cx}{\sigma}\right)\right| \leqq \frac{\varepsilon}{6} \quad \text{and} \quad \left|\Phi\left(\frac{x - 2n^{\frac12}/ck}{\tilde\sigma_k}\right) - \Phi\left(\frac{cx}{\sigma}\right)\right| \leqq \frac{\varepsilon}{6} .$$

Then, for $n \geqq N_2$ (with $k = k(n)$),

$$P\{n^{\frac12}(\hat\theta_n - \theta_0) \leqq x\} \leqq P\left\{|n^{\frac12}(\hat\theta_n{}^k - \theta_0)| \leqq x + \frac{2n^{\frac12}}{ck}\right\} + \frac{\varepsilon}{6}$$

$$\leqq \Phi\left(\frac{x + 2n^{\frac12}/ck}{\tilde\sigma_k}\right) + \frac{\varepsilon}{3} \leqq \Phi\left(\frac{cx}{\sigma}\right) + \frac{\varepsilon}{2}$$

and similarly, $P\{n^{\frac12}(\hat\theta_n - \theta_0) \leqq x\} \geqq \Phi(cx/\sigma) - \varepsilon/2$ and the theorem is proven.

REMARK. It is clear that the conditions for these theorems hold for the minimax Ψ given by (3.11) for any distribution G with a bounded density and giving sufficient probability to a small enough neighborhood of the origin. Furthermore, if G is contaminated by finitely many point masses, there will be at most a finite set of values of ρ for which the conditions of Theorem A.4 do not hold.

Acknowledgments. I would like to thank Larry Brown for an important comment leading to the results of Section 1. I would also like to thank the referees for a number of helpful remarks on the paper.

REFERENCES

[1] ANDREWS, D. F. et al. (1972). *Robust Estimates of Location*. Princeton Univ. Press.

[2] FILIPPOVA, A. A. (1962). Mises theorem of the asymptotic behavior of functionals of empirical distribution functions and statistical applications. *Theor. Probability Appl.* **7** 24-57.

[3] GASTWIRTH, J. L. and RUBIN, H. (1975). The behavior of robust estimators on dependent data. *Ann. Statist.* **3** 1070-1100.

[4] HUBER, P. J. (1964). Robust estimation of a location parameter. *Ann. Math. Statist.* **35** 73-101.

[5] HUBER, P. J. (1972). Robust statistics; a review. *Ann. Math. Statist.* **43** 1041-1067.

[6] STEIN, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **2** 583-602, Univ. of California Press.

[7] STONE, CHARLES (1974). Asymptotic properties of estimators of a location parameter. *Ann. Statist.* **2** 1127-1137.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ILLINOIS
URBANA, ILLINOIS 61801