N° 46 *** 1987

THE MAXIMUM BIAS OF ROBUST CO-VARIANCES

R.A.MARONNA and V.J.YOHAI

notas de matemática

UNIVERSIDAD NACIONAL DE LA PLATA
FACULTAD DE CIENCIAS EXACTAS - DEPARTAMENTO DE MATEMATICA
BIBLIOTECA

N° 46 *** 1987

THE MAXIMUM BIAS OF ROBUST CO-
VARIANCES

R.A.MARONNA and V.J.YOHAI

The Maximum Bias of Robust Covariances

BY RICARDO A. MARONNA AND VICTOR J. YOHAI

ABSTRACT

This paper deals with the maximum asymptotic bias of two classes of robust estimates of the dispersion matrix V of a p-dimensional random vector x, under a contamination model of the form $P = (1-\epsilon)P_0+\epsilon\delta(x$ where P is the distribution of x, P_0 is a spherical distribution, and $\delta(x_0)$ is a point mass at x_0. Estimat... $V_{Q,\alpha}$ of the first class minimize the α quantile of $x'V^{-1}x$ among all symmetric positive-definite matrices V for some $\alpha \in (0,1)$. The "minimum volume ellipsoid" estimator proposed by Rousseeuw belongs to this class with $\alpha = 0.5$. These estimators have breakdown point $\min(\alpha, 1-\alpha)$ for all p. The second class of estimators consist of the M-estimators, from which the seemingly most robust member was chosen; namely the Tyler estimate defined as the solution V_T of $Exx'V_T^{-1}x/x'x = V_T$. This estimator has breakdown point $1/p$. The numerical results show that except for ϵ very close to $1/p$, V_T has in general a smaller maximum bias than $V_{Q,\alpha}$; and that the maximum bias of the latter may be extremely large even for ϵ much smaller than its breakdown point.

Key words and phrases. Robust covariance, maximum bias, M-estimators, high breakdown point estimators

Authors' footnote. Ricardo Maronna is Professor at the Departamento de Matemática, Facultad de Ciencias Exactas, Universidad de La Plata, Calle 50 y 115, 1900 La Plata, Argentina, and Researcher at the C.I.C.P.B.A, La Plata, Argentina.

Victor J. Yohai is Professor at the Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pabellón 1, 1426 Buenos Aires, Argentina, and Researcher at CEMA, Virrey del Pino 3210, 1428 Buenos Aires, Argentina.

1. INTRODUCTION

Let x be a p-dimensional vector with an ellipsoidal distribution, i.e., $x = Ay+\mu$, where A is a nonsingular $p\times p$ matrix, μ is a p-dimensional vector and the random vector y is spherically distributed. Several methods have been proposed to estimate the "dispersion matrix" $V = AA'$ (where $'$ stands for the transpose) and the location vector μ when the data may contain outliers. Among them, M-estimators are defined as solutions of the equations

$$\sum_i w(d_i)(x - \hat{\mu}) = 0 \tag{1-1}$$

$$\sum_i u(d_i^2)(x_i - \hat{\mu})(x_i - \hat{\mu})' = \sum_i v(d_i^2)\hat{V} \tag{1-2}$$

where $d_i^2 = (x_i - \hat{\mu})'\hat{V}^{-1}(x_i - \hat{\mu})$ and u, v, and w are positive functions. Under suitable assumptions on these functions (see Maronna (1976) and Huber (1981)) these estimators are consistent and asymptotically normal. They are affine-invariant and they can be computed numerically by an iterative algorithm. A limit case is the estimator $\hat{V}_T$ studied by Tyler (1987), which has $v = 1$ and $u(t) = p/t$.

One measure of the robustness of an estimator is its asymptotic bias when the distribution P of x is "contaminated" by a fraction of outliers, i.e., the difference between the asymptotic value of the estimator at the "central model" for x, here an ellipsoidal distribution P_0, and the asymptotic value for the distribution $P = (1-\epsilon)P_0 + \epsilon Q$, where Q is an arbitrary distribution. For M-estimators, it has been proved (Maronna (1976), Huber (1981)) that the maximum bias of $\hat{V}$ when μ is known is infinite if $\epsilon > 1/p$, i.e., the breakdown point $\delta^*(\hat{V})$ is always $\leq 1/p$. This fact can be proved by taking Q of the form $\delta(x_0)$, i.e. a point mass concentrated at x_0.

Stahel (1981) and Donoho (1982) defined estimators with high breakdown point for all p, but little is known about them and their numerical calculation seems unfeasible up to now.

A recent attempt to overcome this lack of robustness of $\hat{V}$ is due to Rousseeuw (1986) who proposed to choose the estimators $\hat{\mu}$ and $\hat{V}$ so that ellipsoid $\{x : d^2(x, \hat{\mu}, \hat{V}) \leq 1\}$, where $d^2(x, \mu, V) = (x-\mu)'V^{-1}(x-\mu)$, has minimum volume among all those covering half of the data. This is equivalent to minimize the median of $\{d^2(x_i, \mu, V), \quad i = 1, ..., n\}$ subject to $\det(V) = 1$. He proved that this estimator has breakdown point 0.5 for all p. This proposal can be generalized by considering the α- quantile of d^2 instead of the median, and a similar calculation shows that in that case the breakdown point is $\min(\alpha, 1-\alpha)$. We denote these estimators by $\hat{V}_{Q,\alpha}$ for short.

Davies (1987) proposes a more general class of estimators based on minimizing a robust M-estimate of scale of the $d^2(x_i, \mu, V)$'s, subject to $\det(V) = 1$. This class contains the Q-estimators.

We have computed the maximum asymptotic biases of $\hat{V}_T$ and $V_{Q,\alpha}$ when P_0 is the normal distribution, under the contamination of point masses, and found out that for $\hat{V}_{Q,\alpha}$ it may be large for ϵ substantially smaller than δ^*, while the bias of $\hat{V}_T$ is generally smaller, except for ϵ very close to its breakdown point $1/p$.

In Section 2 we exhibit and comment the numerical results. Section 3 contains the detailed results and their proofs.

2. THE MAXIMUM BIAS

From now on we shall deal only with the estimation of the matrix V, which is the main difficulty, and hence we suppose μ known and equal to 0. We shall take for P_0 the normal distribution. We shall study the asymptotic behavior of the estimators $\hat{V}_T$ and $\hat{V}_{Q,\alpha}$ when the distribution of x is $P = P(\epsilon, x_0) = (1-\epsilon)P_0 + \epsilon\delta(x_0)$. Put $K = |x_0|^2$, where $|x|$ stands for the Euclidean norm of x. Because of the invariance of the estimators, we may assume without loss of generality that $P_0 = N(0, I)$, the unit spherical normal distribution, and that $x_0 = K^{1/2}e_1$ (where $e_1, ..., e_p$ is the canonical basis of R^p).

For a given α, let $q(z)$ stand for the α-quantile of the random variable z under P. Then the asymptotic value of $\hat{V}_{Q,\alpha}$ is the symmetric positive definite matrix $V_{Q,\alpha}$ minimizing $q(x'V^{-1}x)$ under $\det(V) = 1$. The asymptotic value of $\hat{V}_T$ is the solution V_T of the equation

$$pE_P \frac{xx'}{x'V^{-1}x} = V, \tag{2-1}$$

also under the restriction $\det(V) = 1$, which is needed to ensure uniqueness of the solution.

Fix $\epsilon > 0$. Let $V = V(\epsilon, K)$ stand for either $V_{Q,\alpha}$ or V_T under $P(\epsilon, K^{1/2}e_1)$. Then a symmetry argument shows that $V = (v_{ij})$ is diagonal, with elements $v_{11} = b(K)$ (say) and $v_{jj} = a(K)$ for $j > 1$. Since a and b satisfy

$$\det(V) = ba^{p-1} = 1, \tag{2-2}$$

all the information about the bias is contained in $b(K)$, or equivalently in the "condition number" of V : $c(K) = b(K)/a(K) = b(K)^{p/(p-1)}$. We are interested in studying $b^* = \sup_{K>0} b(K)$. Its calculation is described in Theorems 1 and 2 of Section 3.

The numerical values of b^* were computed for several combinations of p and ϵ for V_T and $V_{Q,\alpha}$ with different α's. We report only the values corresponding to $\alpha = 0.5$ and $\alpha = 0.75$, since it was observed that for $\alpha < 0.5$ the bias is larger. Actually it was considered that the maximum condition number $c^* = b^{*p/(1-p)}$ would be more representative. The values of c^* for the three estimates appear in the table respectively under c_T, c_{50} and c_{75}.

To have another measure of "how large" b^* (or c) is, we consider the maximum bias of correlation coefficients. A "robust correlation coefficient" between the coordinates i and j of x, is obtained from the matrix V by $\rho_{ij} = v_{ij}/(v_{ii}v_{jj})^{1/2}$. Assume now that the contamination is located at $x_0 = t(e_1 + e_2)$, and let $K = 2t^2 = |x_0|^2$. Then the distribution of x is obtained from the former case by an orthogonal transformation, and the equivariance of V implies that now V has the form $v_{11} = v_{22} = r$ (say) and $v_{12} = v_{21} = s$, and $v_{jj} = a$ and $v_{ij} = 0$ for $j > 2$ and $i \neq j$, where r and s depend on K and $a = a(K)$ defined above. Since the transformation preserves the trace and the determinant, it follows that $\mathrm{tr}(V) = 2r + (p-2)a = b + (p-1)a$, where $b = b(K)$ and $\det(V) = (r^2 - s^2)a^{p-2} = 1$. Hence $a + b = 2r$ and $r^2 - s^2 = ab$. Thus the correlation $\rho(K) = \rho_{12} = s/r$ satisfies $\rho(K)^2 = (r^2 - ab)/r^2 = 1 - 4c(K)/(1 + c(K))^2$, so that maximizing ρ is equivalent to maximizing c. The maximum value of ρ for the three estimators, corresponding to putting $c = c^*$ above, are also shown in the table respectively under ρ_T, ρ_{50} and ρ_{75}

Table about here

An analysis of the table shows that $\hat{V}_{Q,\alpha}$ may have a very large bias for ϵ much smaller than its breakdown point of 0.50 or 0.25, and $\hat{V}_T$ behaves better for low ϵ and not very large p $(p \leq 5)$.

Tyler has given an algorithm for the calculation of $\hat{V}_T$ which converges independently of the starting matrix. By the contrary the computation of $\hat{V}_{Q,\alpha}$ requires the absolute minimum of a non convex function, and therefore requires a good starting point to yield reliable results. Thus the actual behavior of the latter will depend on the starting matrix, and thus their advantages with respect to the former may be smaller than the results of Table 1 suggest.

The choice of the Tyler estimator within the class of M-estimators and of the Q-estimators among the class of S-estimators was suggested by te results of Martin, Yohai and Zamar (1987) who show that for regression, the estimators analogous to them minimize the maximum asymptotic bias for a given ϵ among the classes of M- and of S-estimators. We suspect that similar results should hold for the robust covariance problem.

3. MAIN RESULTS AND PROOFS

In this section we state and prove the results which led to the numerical values of Table 1.

We consider first the bias of $\hat{V}_{Q,\alpha}$. By the reasoning above (2-2) we may from the start restrict ourselves to matrices V of the form given there.

Let $\epsilon < 0.5$ and $\epsilon < \alpha < 1-\epsilon$ (the dependency of all values on ϵ and α will in general not be shown explicitly). Let P_0 be a distribution with a density $f(x)$ which is a decreasing function of $|x|$. For $b > 0$ define the random variable $s(b) = x_1^2/b + \sum_{j=2}^{p} x_j^2/a$ where $a = b^{-1/(p-1)}$. Define $q(K,b)$ as the α-quantile of $s(b)$ when x has the distribution $P_K = (1-\epsilon)P_0 + \epsilon\delta(K^{1/2}e_1)$, $b(K)$ as the value of b which minimizes $q(K,b)$ and F_b as the distribution function of $s(b)$ under P_0.

THEOREM 1. *Define K_1, K_0 and b^* by $F_1(K_1) = \alpha/(1-\epsilon)$, $F_{b^*}(K_1) = (\alpha-\epsilon)/(1-\epsilon)$, and $K_0 = b^* K_1$. Then $b^* = \sup_{K>0} b(K)$, and furthermore $b(K)$ is continuous and nondecreasing for $K < K_0$ and $b(K) = 1$ for $K > K_0$ (and hence $\hat{V}_Q$ behaves like a redescending estimator).*

To prove this theorem, an auxiliary result is needed:

LEMMA. *Let the random vector x have a density $f(x)$ which is a decreasing function of $|x|$. Let $1 < m < p$ and define the random variables $u = u(x) = \sum_{i=1}^{m} x_i^2$ and $v = v(x) = \sum_{i=m+1}^{p} x_i^2$. For $b > 0$ let F_b be the distribution function of $s(b) = u/b + v/a$ where $b^m a^{p-m} = 1$. Then for each t, $F_b(t)$ is a decreasing function of b for $b \geq 1$ and is increasing for $b \leq 1$.*

PROOF: We shall prove that if $1 \leq b < b'$ then for all $t > 0$: $P(s(b) \leq t) > P(s(b') \leq t)$. Let $A = \{x \in R^p : u(x)/b + v(x)/a \leq t\}$ and $B = \{x \in R^p : u(x)/b' + v(x)/a' \leq t\}$, where

$$b^m a^{p-m} = b'^m a'^{p-m} = 1. \tag{3-1}$$

We want to prove that $\int_A f > \int_B f$. This is equivalent to proving $\int_{A-B} f > \int_{B-A} f$. Now (3-1) implies that (since A and B are ellipsoids) $m(A) = m(B)$, where m is the Lebesgue measure. Hence $m(B-A) = m(A-B)$. Thus it will suffice to show that if $x \in A - B$ and $y \in B - A$ then $f(x) > f(y)$.

Let (u_0, v_0) be the solution of the linear system

$$\frac{u_0}{b} + \frac{v_0}{a} = t, \qquad \frac{u_0}{b'} + \frac{v_0}{a'} = t, \tag{3-2}$$

and put $w_0 = u_0 + v_0$. Then it will be shown that $x \in A - B$ and $y \in B - A$ implies $|x|^2 < w_0 < |y|^2$. To prove the first inequality, recall that $|x|^2 = u(x) + v(x)$. Thus it will be shown that for any $u, v \in R$

$$\frac{u}{b} + \frac{v}{a} \leq t \quad \text{and} \quad \frac{u}{b'} + \frac{v}{a'} > t \quad \text{imply} \quad u + v < w_0. \tag{3-3}$$

Recall $a' < a < b < b'$. Note that u_0 and v_0 can be calculated explicitly from (3.2). Multiplying the first inequality in (3-3) by a and the second by $-a'$ and adding both, yields $u < t(a - a')/(a/b - a'/b') = u_0$. The first inequality in (3-3) together with the first equation in (3-2) yield $v \leq v_0 + (a/b)(u_0 - u)$. Hence $u + v = v + (u - u_0) + u_0 < v_0 + u_0 + (u_0 - u)((a/b) - 1) < w_0$. This proves $|x|^2 < w_0$, and the remaining inequality follows likewise. Now since f is decreasing, $|x| < |y|$ implies $f(x) > f(y)$ as stated. The case $b \leq 1$ follows easily from this one.

PROOF OF THEOREM 1: We shall apply the Lemma to the case $m = 1$. Thus the function $F_b(t)$ is continuous and increasing in t, and is continuous and decreasing (increasing) in b for $b \geq 1$ ($b \leq 1$). Define $k(b) = F_b^{-1}(\alpha/(1-\epsilon))$ and $j(b) = F_b^{-1}((\alpha - \epsilon)/(1-\epsilon))$. Thus $k(b)$ and $j(b)$ are continuous and increasing (decreasing) for $b \geq 1$ ($b \leq 1$), and $j(b) < k(b)$ for all b. Let $K_1 = k(1)$ and $J_1 = j(1)$.

We shall first calculate $q(K, b)$. If x has the distribution P_K (recall the definitions above the Theorem 1) then $x(b)$ has distribution function $P(x(b) \leq t) = (1-\epsilon)F_b(t) + \epsilon I(t > K/b)$ where I is the indicator function. The α-quantile $q = q(K, b)$ of this distribution verifies:

$$q = k(b) \qquad \text{for} \qquad bk(b) \leq K \tag{3-4}$$

$$q = \frac{K}{b} \qquad \text{for} \qquad bk(b) > K \geq bj(b) \tag{3-5}$$

$$q = j(b) \qquad \text{for} \qquad bj(b) > K. \tag{3-6}$$

Now we minimize q as a function of b. Put $q_0(K) = \min_{b>0} q(K, b)$ and let $b(K)$ be the value of b which yields the minimum. Let $b_0(K)$ and $b_1(K)$ be respectively the solutions of $bj(b) = K$ and of $bk(b) = K$ with $b \geq 1$. Thus $b_0 > b_1$ and both are increasing functions. We consider three cases to study q as a function of b;

(a) $K > K_1$. It follows from (3-4) that q decreases for $b \leq 1$ and increases for $1 \leq b \leq b_0(K)$; by (3-5) it decreases for $b_1(K) \leq b \leq b_0(K)$, and by (3-6) it increases for $b > b_0(K)$. Thus q has two local minima at $b = 1$ and $b = b_0(K)$, and hence if $j(b_0(K)) > K_1$ then $q_0(K) = K_1$ and $b(K) = 1$; if $j(b_0(K)) < K_1$ then $q_0(K) = j(b_0(K))$ and $b(K) = b_0(K)$; and if $j(b_0(K)) = K_1$ then $q_0(K) = K_1$ and $b(K) \in [1, b_0(K)]$ (i.e., the minimum is attained at these two points).

(b) $K_1 \geq K > J_1$. Note that $b = 1$ belongs to (3-5). An analysis similar to (a), but simpler, shows that $b(K) = b_0(K)$ and $q_0(K) = j(b_0(K))$.

(c) $J_1 \geq K$. A similar analysis yields $b(K) = 1$ and $q_0(K) = J_1$.

Let K_0 satisfy $j(b_0(K_0)) = K_1$. Then from (a), (b) and (c) it follows that $b(K) = 1$ for $K \leq J_1$, $b(K) = b_0(K)$ for $J \leq K < K_0$, $b(K) = 1$ for $K > K_0$; for $K = K_0$ it takes on the values 1 and $b_0(K_0)$. Since b_0 is increasing, it follows that the supremum of $b(K)$ is $b_0(K_0) = b^*$. The definition of the functions j and b_0 imply that $K_0 = b^* K_1$.

The calculation of the bias of the Tyler estimate is easier. We have the following Theorem.

THEOREM 2. *Let P_0 be any spherical distribution such that $P_0(z = 0) = 0$. Let $\epsilon < 1/p$, and let $o = o(K)$ be the condition number of V_T under P_K as defined above. Then for all $K > 0$, o is the solution of*

$$\frac{(1/p) - \epsilon}{1 - \epsilon} = E \frac{t}{t + o(1 - t)}, \tag{3-7}$$

~~*where t has a Beta distribution with degrees of freedom*~~ 1 *and* $p - 1$.

PROOF: Put in (2-1) $P = P_K$, take V of the form described above (2-2), and pre- and post- multiply respectively by e_1' and e_1. It follows that

$$(1 - \epsilon) E_{P_0} \frac{u}{u/b + v/a} + \epsilon \frac{K}{K/b} = \frac{b}{p},$$

where u and v are defined in the statement of the Lemma above , with $m = 1$. Putting $o = b/a$ and $t = u/(u + v)$ yields (3-7). The distribution of t is the same for any distribution of $|z|$ which gives null mass to the origin, and hence is Beta.

Numerical computations. The distribution function F_b was evaluated as

$$F_b(z) = \int_0^1 H\left(\frac{z}{b + t(a - b)}\right) g(t) dt,$$

where H is the distribution function of the chi-squared distribution with p degrees of freedom, and g is the density of the Beta distribution with degrees of freedom 1 and $p - 1$. The function H was evaluated by means of the subroutine CDTR of the IBM Scientific Subroutine Package. The change of variable $t' = (1 - t^2)^2$ was used to avoid the singularities of the integrand at the end points; and the integral was evaluated by means of 32-point Gauss quadrature implemented in the double precision subroutine DQG32 of the IBM Package. The equations needed to compute b were solved using the subroutine RTMI in the same package. The same methods were used for solving (3-7). An IBM 4331 computer was used.

REFERENCES

Donoho,D.L. (1982), "Breakdown properties of multivariate location estimators", Ph.D. qualifying paper, Harvard University.

Davies, P.L. (1986), "Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices", unpublished manuscript.

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

Maronna,R (1976), "Robust M-estimators of multivariate location and scatter", *Annals of Statistics*, 4, 51-67.

Martin,R.D., Yohai,V.J. and Zamar,R. (1987), "Minimax bias robust estimation of regression", unpublished manuscript.

Rousseeuw,P.J., (1985) "Multivariate estimation with hight breakdown point", in: *Mathematical Statistics and Applications*, edited by W.Grossmann, G.Pflug, I.Vincze and W.Wertz, Reidel Publishing Company, 283-297.

Stahel,W. (1981), "Breakdown of covariance estimators", Research Report No. 31, Fachgruppe für Statistik, ETH, Zurich.

Tyler,D.E. (1987),"A distribution free M-estimator of multivariate scatter, *Annals of Statistics*,15.

TABLE

MAXIMUM BIAS OF V_T AND $V_{Q,\alpha}$

p	ϵ	c_T	c_{50}	c_{75}	ρ_T	ρ_{50}	ρ_{75}
2	0.05	1.2	6.3	3.8	0.10	0.73	0.58
	0.10	1.6	14.2	7.1	0.22	0.87	0.75
	0.20	2.8	58.3	27.5	0.47	0.97	0.93
	0.25	4.0	129.0	∞	0.60	0.98	1.00
3	0.05	1.3	4.6	3.2	0.13	0.65	0.52
	0.10	1.8	9.5	5.5	0.28	0.81	0.69
	0.20	4.3	32.0	15.9	0.62	0.94	0.88
	0.25	8.7	60.5	∞	0.79	0.97	1.00
4	0.05	1.4	4.1	3.0	0.16	0.61	0.51
	0.10	2.1	8.1	5.1	0.35	0.78	0.67
	0.20	9.0	25.8	13.6	0.80	0.92	0.86
	0.24	59.6	41.5	26.9	0.97	0.95	0.93
	0.25	∞	46.9	∞	1.00	0.96	1.00
5	0.05	1.5	3.9	3.0	0.19	0.59	0.50
	0.10	2.5	7.5	5.0	0.43	0.76	0.67
	0.15	6.0	13.2	7.8	0.71	0.86	0.77
	0.20	∞	23.1	12.8	1.00	0.92	0.85
10	0.05	2.2	3.6	3.1	0.38	0.57	0.51
	0.10	∞	6.7	5.0	1.00	0.74	0.67
	0.20	∞	19.3	12.5	1.00	0.90	0.85
15	0.04	2.7	3.2	2.8	0.46	0.52	0.48
	0.05	4.5	3.7	3.2	0.63	0.57	0.52
	0.10	∞	6.7	5.3	1.00	0.74	0.68
	0.20	∞	18.9	13.1	1.00	0.90	0.86
20	0.03	2.7	2.8	2.6	0.45	0.47	0.44
	0.05	∞	3.7	3.4	1.00	0.58	0.54
	0.10	∞	6.8	5.6	1.00	0.74	0.70
	0.20	∞	19.0	13.3	1.00	0.90	0.87

UNIVERSIDAD NACIONAL DE LA PLATA

Facultad de Ciencias Exactas - Departamento de Matemática

Casilla de correo 172 - 1900 La Plata - Argentina