The Annals of Statistics
1977, Vol. 5, No. 4, 779–786

A LOCATION ESTIMATOR BASED ON A *U*-STATISTIC

BY J. S. MARITZ, MARGARET WU AND R. G. STAUDTE, JR.

C.S.I.R.O. Division of Mathematics and Statistics, Monash University and La Trobe University

Let $X_1, \cdots, X_n$ be i.i.d. F, and estimate the median of F by the median T_β of $\beta X_i + (1-\beta)X_j$, $i \neq j$, where β is a fixed positive constant. Then T_β is the solution of a U-statistic equation from which its asymptotic normality is readily derived. The asymptotic relative efficiency of T_β is computed for a few cdfs F and seen to be reasonably high for unintuitive choices such as $\beta = .9$, $\beta = 2$, and also to be remarkably constant for $\beta > 1$. Moreover, the influence curves and breakdown points of $\{T_\beta : \beta > 0\}$ are derived and indicate that the good robustness properties of the Hodges–Lehmann estimator ($\beta = \frac{1}{2}$) are shared by the entire class.

Monte Carlo estimates of the variance of T_β for sample sizes $n = 10$, 20, and 40 indicate that some of these estimators perform as well as those discussed in the Princeton Robustness Study when the underlying F is double-exponential or Cauchy.

1. Introduction. Given $X_1, \cdots, X_n$ i.i.d. F, consider estimating the median of F by the median of $\{\beta X_i + (1-\beta)X_j : 1 \leqq i \neq j \leqq n\}$, where β is a fixed positive constant. The case $\beta = 1$ yields the sample median; $\beta = \frac{1}{2}$ yields the Hodges–Lehmann estimator. The asymptotic efficiency and robustness of these estimators are studied for various choices of F and β. These estimators arise as solutions to a U-statistic equation

$$\sum_{i<j} \Phi(X_i - \theta, X_j - \theta) = 0 \tag{1.1}$$

where $\Phi(x, y) = \frac{1}{2}[\mathrm{sgn}\,(\beta x + (1-\beta)y) + \mathrm{sgn}\,(\beta y + (1-\beta)x)]$ is the defining symmetric kernel. The asymptotic normality of such estimators follows from the results of Hoeffding (1948); see Section 3. The robustness properties are more easily established by consideration of the solution to (1.1) in which the $i = j$ pairs (i.e., the original observations) are included. This estimator may be expressed $T_\beta(F_n)$, where F_n is the empirical cdf and $T_\beta(F)$ is the functional defined for arbitrary cdf F by the solution of the equation

$$\lambda_F(\theta) \equiv \textstyle\iint \Phi(x-\theta, y-\theta)F(dx)F(dy) = 0\,. \tag{1.2}$$

(To ensure the existence of a unique solution to (1.2) for arbitrary F, we define $T_\beta(F)$ to be the midpoint of the interval of θ values for which $\lambda_F(\theta + \eta) \leqq 0$ and $\lambda_F(\theta - \eta) \geqq 0$ for all $\eta > 0$.) It is easily seen that $T_\beta(F)$ is the median of the distribution of $\beta X + (1-\beta)Y$, where X, Y are i.i.d. F. When F is symmetric $T_\beta(F) =$ median of F, so $T_\beta(F_n)$ is a reasonable estimator of the median. When

Received December 1975; revised December 1976.

AMS 1970 *subject classifications.* Primary 62G05; Secondary 62G35.

Key words and phrases. Location estimator, U-statistic, robustness, influence curve, breakdown point, asymptotic relative efficiency, Hodges–Lehmann estimator.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics.*
www.jstor.org

F is asymmetric, $T_\beta(F)$ will in general differ from the median and then $T_\beta(F_n)$ will not even be a consistent estimator of it. However, if $T_\beta(F)$ and the median are close, $T_\beta(F_n)$ may be a reasonable estimate for small samples; this possibility is not further explored here.

In Section 4 of the asymptotic variances of T_β are tabled for comparison with those of maximum likelihood estimates in three cases of normal, Cauchy and double-exponential densities. It will be seen that values of $\beta > 1$ are not unreasonable choices for the latter two densities and that $\beta = .9$ maximizes the minimum efficiency over the three densities. Monte Carlo estimates of the variance of $T_\beta(F_n)$ are tabled for $n = 10, 20$, and 40 for comparison with estimators considered by Crow and Siddiqui (1967) and those of the Princeton study (1972).

The estimators discussed here fall within the large class of M_2-estimators introduced by Huber (1964). Test of symmetry based on U-statistics are discussed by M. K. Gupta (1967), but there is little overlap with the material presented here.

2. Robustness properties of the estimator T_β. It is convenient to rewrite $\lambda_F(\theta) = E_F k_F(Y, \theta)$, where

$$k_F(y, \theta) = 1 - F\left(\frac{\theta - \beta y}{1 - \beta}\right) - F\left(\frac{\theta - (1 - \beta)y}{\beta}\right), \qquad 0 < \beta < 1 \tag{2.1}$$

$$= F\left(\frac{\theta - \beta y}{1 - \beta}\right) - F\left(\frac{\theta - (1 - \beta)y}{\beta}\right), \qquad \beta > 1 .$$

Note that $k_F(y, \theta)$ is not a function of $y - \theta$, even though $T_\beta(F)$ is location invariant. Some of the robutsness properties (Hampel's definitions [4]) of the estimator are obtained in the following proposition.

PROPOSITION 2.1. *Let F be a continuous distribution for which there is a unique solution to* (1.2). *Then*

(i) *T_β is continuous at F with respect to the topology of weak convergence*;
(ii) *the breakdown point $\delta_\beta{}^*$ of T_β is* $1 - 2^{-\frac{1}{2}} \approx .29$ *for all* $0 < \beta \neq 1$;
(iii) *if F has square integrable density f which is positive and continuous on an open interval containing $T_\beta(F)$, then the influence curve of T_β at F is*

$$\Omega_\beta(x) = \frac{-2k_F(x, T_\beta(F))}{E_F k_F{}'(X, T_\beta(F))} \tag{2.2}$$

where $k_F{}'(x, \theta) = (\partial/\partial\theta)k_F(x, \theta)$; *and*

(iv) *under the conditions stated in* (iii), *the "gross error sensitivity"* $\gamma_\beta{}^* = \sup_x |\Omega_\beta(x)| \leqq 1/E_F k_F{}'(X, T_\beta(F))$ *with equality if* $0 < \beta < 1$.

The boundedness and continuity of Ω_β indicate that T_β will be rather insensitive to outliers and local contamination. When F is symmetric, (2.2) reduces to

$$\Omega_\beta(x) = \frac{\beta[F(\beta x/|1 - \beta|) - F(-((1 - \beta)/\beta)x)]}{\int f(((1 - \beta)/\beta)y)f(y)\, dy} . \tag{2.3}$$

For $0 < \beta < 1$, $\Omega_\beta(x)$ is odd and monotone increasing as well as bounded so outliers tend to have the same effect.

When $\beta > 1$, $|\Omega_\beta(x)|$ tends to zero as $|x|$ increases so that outliers have a diminishing effect as their distance from the parameter increases. The examples in Section 4 indicate that the asymptotic variance of T_β (which will be seen to equal $E_F\Omega_\beta^2(X)$) is remarkably constant for values of $\beta > 2$. Thus we point out that

$$\lim_{\beta\to\infty} \Omega_\beta(x) = \frac{2xf(x)}{\int f^2(y)\,dy}. \tag{2.4}$$

The case $\beta = 1$ yields the sample median, which has already been discussed in the literature [4]. By way of comparison, T_1 is continuous at F if and only if $\{F^{-1}(\frac{1}{2})\}$ is a singleton; $\delta_1{}^* = \frac{1}{2}$; $\gamma_1{}^* = 1/2f(0)$; and $\Omega_1(x) = (\text{sgn}\,(x))/2f(0)$. From (2.3) we see that for continuous f, $\lim_{\beta\to 1} \Omega_\beta(x) = \Omega_1(x)$.

PROOF OF PROPOSITION 2.1 (for $0 < \beta < 1$). (i) Roughly speaking, if F, G are close in the topology of weak convergence, then k_F, k_G are uniformly close and consequently $\lambda_F(\theta)$, $\lambda_G(\theta)$ are close for all θ, including $\theta = T_\beta(F)$, $T_\beta(G)$. More precisely, let $\lambda(F, G)$ and $\pi(F, G)$ denote respectively Lévy and Prohorov metrics (each of which generates the topology). Then $\pi(F, G) < \delta$ implies $\lambda(F, G) < \delta$, which in turn implies

$$|k_G(y, \theta) - k_F(y, \theta)| < 2(\delta + \alpha_F(\delta))$$

where $\alpha_F(\delta) \equiv \sup_x |F(x + \delta) - F(x)|$ is the modulus of continuity of F. It follows that

$$\begin{aligned} |\lambda_G(\theta) - \lambda_F(\theta)| &= |\textstyle\int k_G(y, \theta)G(dy) - \int k_F(y, \theta)F(dy)| \\ &\leqq 2(\delta + \alpha_F(\delta)) + |\textstyle\int k_F(y, \theta)G(dy) - \int k_F(y, \theta)F(dy)|\,. \end{aligned} \tag{2.4}$$

By result (a) of Section II.8, Oaten (1972), $\pi(F, G) < \delta$ implies that the quantity in absolute values is bounded by $2\delta + \alpha_{k_F(\cdot,\theta)}(\delta)$, which is independent of θ. Hence the left-hand side of (2.4) converges to G uniformly in θ as $\delta \to 0$. When the solution to (1.2) is unique any solution to $\lambda_G(\theta) = 0$ will clearly converge to it as $\pi(F, G) \to 0$.

PROOF OF (ii). By definition, $\delta_\beta{}^* = \sup\{\delta : \sup_{\pi(F,G)<\delta} |T_\beta(G)| < \infty\}$.

We first show that $T_\beta(G)$ is bounded on $\{G : \pi(F, G) < \delta\}$ only if $\delta < 1 - 2^{-\frac{1}{2}}$. To this end define

$$F_{x,\delta} = (1 - \delta)F + \delta\varepsilon_x \tag{2.5}$$

where ε_x places mass 1 on x. Then writing out $\int k_{x,\delta}\,dF_{x,\delta}$ we find that $T_\beta(F_{x,\delta})$ is the solution of

$$\int k(y, \theta)F(dy) = \frac{-\delta}{1-\delta}\left[2k_F(x, \theta) + \left(\frac{\theta}{1-\theta}\right)\text{sgn}\,(x - \theta)\right]. \tag{2.6}$$

The left-hand side of (2.6) is monotonically decreasing from 1 to -1 while the right-hand side is monotonically increasing from $-\delta(2 - \delta)/(1 - \delta)^2$ to

$\delta(2-\delta)/(1-\delta)^2$. The solution(s) of (2.6) are bounded as $|x| \to \infty$ if and only if $\delta(2-\delta)/(1-\delta)^2 < 1$. Hence $\delta_\beta^* \leqq 1 - 2^{-\frac{1}{2}}$.

To prove the reverse inequality, fix $\pi(F, G) < \delta < 1 - 2^{-\frac{1}{2}}$. We will show that solutions to $\lambda_G(\theta) = 0$ are bounded by a constant independent of G. Again using the fact that $\pi(F, G) \geqq \lambda(F, G)$, we have

$$\int G\left(\frac{\theta - \beta y}{1-\beta}\right) G(dy) \geqq \int_{-\infty}^{\theta-(1-\beta)x_\delta/\beta} \left[F\left(\frac{\theta - \beta y}{\beta} - \delta\right) - \delta\right] G(dy)$$

where $x_\delta = F^{-1}(\delta) + \delta$. The integrand is nonincreasing in y, and G is stochastically smaller than F_θ on the range of integration, where

$$\begin{aligned} F_\theta(y) &= 0 & & y < x_\delta \\ &= F(y-\delta) - \delta & & x_\delta \leqq y < \frac{\theta - (1-\beta)x_\delta}{\beta} \\ &= 1 & & y \geqq \frac{\theta - (1-\beta)x_\delta}{\beta}. \end{aligned}$$

Hence

$$\int G\left(\frac{\theta - \beta y}{1-\beta}\right) G(dy) \geqq \int_{x_\delta}^{\theta-(1-\beta)x_\delta/\beta} \left[F\left(\frac{\theta - \beta y}{1-\beta} - \delta\right) - \delta\right] F_\theta(dy),$$

which converges to $(1-\delta)^2$ as $\theta \to +\infty$. The same computation on $\int (G(\theta - (1-\beta)y)/\beta) G(dy)$ shows $\lambda_G(\theta)$ is bounded above by a function of F and θ which tends to $1 - 2(1-\delta)^2 < 0$ as θ goes to $+\infty$. Thus any solution $T_\beta(G)$ is bounded above by a constant independent of G. Similarly $T_\beta(G)$ is bounded below, and $\delta_\beta^* \geqq 1 - 2^{-\frac{1}{2}}$.

PROOF OF (iii). Define $\lambda_{x,\delta}(\theta) = \int k_{x,\delta}(y, \theta) F_{x,\delta}(dy)$ where $k_{x,\delta}$ is the kernel corresponding to (2.5). Then $\lambda_{x,\delta}(\theta) = (1-\delta)^2 \int k_F(y, \theta) F(dy) + 2\delta(1-\delta) k_F(x, \theta) + \delta^2 \operatorname{sgn}(x - \theta)$ has strictly negative derivative for all θ near $T_\beta(F)$ under the given conditions on F. Moreover the solution of $\lambda_{x,\delta}(\theta) = 0$, namely $\theta = T_\beta(F_{x,\delta})$, is near $T_\beta(F)$ for all sufficiently small δ. Hence by the mean value theorem

$$T_\beta(F_{x,\delta}) = -\lambda_{x,\delta}(0)/\lambda'_{x,\delta}(\gamma T_\beta(F_{x,\delta}))$$

where $\gamma = \gamma(x, \delta)$ satisfies $0 \leqq \gamma \leqq 1$. The result now follows easily from the definition

$$\Omega_\beta(x) = \lim_{\delta \downarrow 0} \frac{T_\beta(F_{x,\delta}) - T_\beta(F)}{\delta}.$$

The arguments for (i)—(iii) when $\beta > 1$ are analogous to those just given for $0 < \beta < 1$ and are thus omitted. The proof of (iv) is trivial.

3. Consistency and asymptotic normality of $T_\beta(F_n)$. The sequence of estimators $\{T_\beta(F_n)\}$ is strongly consistent for $T_\beta(F)$ whenever the latter is the unique solution to (1.2). This follows from the continuity of T_β at F (Proposition 2.1 (i)) and the corollary to Glivenko–Cantelli that $\lambda(F_n, F)$ converges to 0 almost surely. The asymptotic normality of $\{T_\beta(F_n)\}$ is obtained in the following proposition.

PROPOSITION 3.1. *Fix* $0 < \beta \neq 1$. *Assume F has square-integrable density f and that* (1.2) *has unique solution* $T_\beta(F)$. *Then* $n^{\frac{1}{2}}[T_\beta(F_n) - T_\beta(F)]$ *is asymptotically normal with parameters* $(0, E_F\Omega_\beta^2(X))$, *where* $\Omega_\beta(x)$ *is given by* (2.2).

PROOF. Assume without loss of generality that $T_\beta(F) = 0$. Let $\underline{T}_\beta(F_n)$ and $\bar{T}_\beta(F_n)$ denote the infimum and supremum of the interval of solutions to $\lambda_{F_n}(\theta) = 0$. We will show that $n^{\frac{1}{2}}\bar{T}_\beta(F_n)$ is asymptotically normal $(0, E_F\Omega_\beta^2(X))$; an identical argument for $n^{\frac{1}{2}}\underline{T}_\beta(F_n)$ leads to the same asymptotic distribution, completing the proof.

First note that $P\{n^{\frac{1}{2}}\bar{T}_\beta(F_n) < x\} = P\{\lambda_{F_n}(x/n^{\frac{1}{2}}) < 0\}$. It follows from the results in Hoeffding (1948) that $n^{\frac{1}{2}}[\lambda_{F_n}(x/n^{\frac{1}{2}}) - \lambda_F(x/n^{\frac{1}{2}})]$ is asymptotically normal with parameters $(0, 4\,\mathrm{Var}_F\, k_F(X, 0))$. Moreover, $n^{\frac{1}{2}}\lambda_F(x/n^{\frac{1}{2}}) = \int n^{\frac{1}{2}}[k_F(y, x/n^{\frac{1}{2}}) - k_F(y, 0)]F(dy)$ converges to $xE_F\, k_F'(X, 0)$ by the dominated convergence theorem (where again, $k_F'(y, \theta) = (\partial/\partial\theta)k_F(y, \theta)$). Therefore $P\{n^{\frac{1}{2}}\bar{T}_\beta(F_n) < x\}$ converges to $\Phi(xE_F[k_F'(X, 0)]/2(\mathrm{Var}_F\, k_F(X, 0))^{\frac{1}{2}})$, and reference to (2.2) completes the argument.

4. Examples of estimator efficiency. The estimators defined by T_β have been shown to possess good robustness properties and it is natural to ask what premium is being paid in efficiency. The asymptotic relative efficiency of T_β relative to maximum likelihood estimators is sketched in Figure 4.1 as a function of β

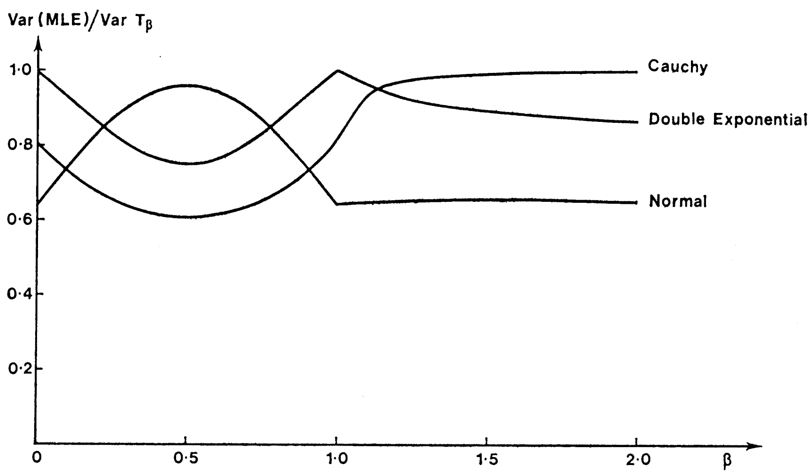

FIG. 4.1.

for the three cases (a) normal, (b) double-exponential, and (c) Cauchy. The Hodges–Lehmann estimator $T_{.5}$ is clearly the most efficient of this class for normal data; the median T_1 is the most efficient in the double-exponential case; and for the Cauchy distribution T_β has efficiency approaching one as β increases without bound.

A value of β near .9 maximizes the minimum efficiency over the three densities. This maximin efficiency is approximately .73 which should be compared to the value .82 obtainable by using the best trimmed mean (Crow and Siddiqui (1967)). On the other hand, if only the wider tailed distributions are considered, the best trimmed mean has maximin efficiency .88 while $T_{1.1}$ attains .94.

In order to gain some insight into the finite sample behavior of T_β Monte Carlo estimates of the variances were found for sample sizes $n = 10, 20$, and 40 and selected values of β. These are listed in Table 4.2 together with the asymptotic variances.

TABLE 4.2
Estimates of $n \times \text{Var}\, T_\beta(F_n)$

n	β								
	.5	.6	.7	.8	.9	1.0	2.0	4.0	6.0
(a) *Normal*									
10	1.06	1.07	1.11	1.20	1.30	1.37 (1.38)*	1.55	1.62	1.65
20	1.06	1.06	1.11	1.20	1.34	1.46 (1.47)*	1.54	1.63	1.70
40	1.05	1.06	1.10	1.20	1.34	1.52	1.55	1.57	1.63
∞	1.05	1.06	1.10	1.20	1.35	1.57	1.53	1.54	1.54
(b) *Double-Exponential*									
10	1.57	1.54	1.46	1.41	1.40	1.45 (1.45)	1.55	1.66	1.69
20	1.54	1.50	1.37	1.32	1.25	1.37 (1.33)	1.42	1.51	1.51
40	1.40	1.36	1.34	1.23	1.17	1.23	1.25	1.33	1.38
∞	1.33	1.31	1.25	1.16	1.07	1.00	1.16	1.18	1.18
(c) *Cauchy*									
10	11.2	7.8	5.9	5.8	4.5	3.33 (3.36)	3.15	3.50	3.50
20	4.5	4.5	4.1	3.8	3.3	2.78 (2.79)	2.50	2.70	2.85
40	3.8	3.7	3.5	3.4	3.1	2.66	2.18	2.39	2.41
∞	3.29	3.25	3.14	2.96	2.73	2.46	2.00	2.00	2.00

* Values in parentheses are exact to two decimal places.

The normal variance estimates were obtained indirectly (see Hodges (1967)) through estimates of Var $(T_\beta - \bar{X})$ based on 4000 samples. Double-exponential estimates were based on 5000 samples; Cauchy estimates on 5000 antithetic samples. The estimated standard error divided by the estimate was calculated for each table entry and found to be largest for a sample of size 10 from the Cauchy distribution and $\beta < 1$, where it was approximately .05. In all other cases it was less than .025 and for most estimates less than .01. Several remarks are in order.

1. The asymptotically most efficient member of this class for normal data, the Hodges–Lehmann estimator, appears to retain its primary position for finite samples.

2. For small samples of exponential data $T_{.9}$ appears to be more efficient than

the asymptotically efficient median. In fact the lowest value of $n \times$ variance for $n = 20$ among the 65 estimators considered in the Princeton study (1971) is 1.29 compared with 1.25 for $T_{.9}$.

3. For samples of size $n = 10$, 20, and 40 from the Cauchy distribution, T_2 does about as well as any of the Princeton study estimators. As mentioned previously, larger choices of β will yield *asymptotic* efficiency as close to 1 as desired. On the other hand, if we fix n and let β tend to infinity, $T_\beta(F_n)$ becomes $T_1(F_n)$, the sample median, which is not efficient in this context. The apparent conflict is resolved by letting $\beta = \beta_n$ approach infinity with n. Then the statistic defining $T_{\beta_n}(F_n)$ is of the form

$$n^2\lambda_{F_n}(\theta) = 2 \sum\sum_{i<j} \Phi_{\beta_n}(X_i - \theta, X_j - \theta) + \sum_i \operatorname{sgn}(X_i - \theta) . \tag{4.1}$$

When β_n grows faster than n, the first term in (4.1) is of smaller order than the second, and the latter sign test statistic leads to the sample median. When β_n is of smaller order than n, the second term is negligible relative to the first. Moreover, the projection of the first term onto the linear space spanned by functions of the form $\sum_i h_i(X_i - \theta_n)$ (as required in the proof of asymptotic normality), is $2 \sum_i k_F(X_i, \theta_n)$ with k_F given by (2.1). As $\beta_n \to \infty$, this is equivalent to $(1/\beta_n) \sum_i(X_i - \theta_n)f(X_i - \theta_n)$ which is the statistic defining the maximum likelihood estimator for the Cauchy distribution. Thus for β_n growing to infinity at any rate less than n, $T_{\beta_n}(F_n)$ will be asymptotically efficient for the Cauchy distribution. The tabled values show that the optimal β_n for a given n may actually be quite small; e.g., the optimal β_{40} is about 2.

Acknowledgment. The authors wish to thank Dr. Bruce M. Brown of La Trobe University for helpful suggestions.

REFERENCES

[1] ANDREWS, et al. (1972). *Robust Estimates of Location; Survey and Advances.* Princeton University Press.

[2] CROW, E. L. and SIDDIQUI, M. M. (1967). Robust estimation of location. *J. Amer. Statist. Assoc.* **62** 353–389.

[3] GUPTA, M. K. (1967). Asymptotically nonparametric tests of symmetry. *Ann. Math. Statist.* **38** 849–866.

[4a] HAMPEL, F. R. (1971). A general qualitative definition of robustness. *Ann. Math. Statist.* **42** 1887–1896.

[4b] HAMPEL, F. R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc.* **69** 383–393.

[5] HODGES, J. L. (1967). Efficiency in normal samples and tolerance of extreme values for some estimates of location. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1** 163–186, Univ. of California Press.

[6] HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19** 293–325.

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

[8] OATEN, A. (1972). Approximation to Bayes risk in compound decision problems. *Ann. Math. Statist.* **43** 1164–1184.

J. S. MARITZ
C.S.I.R.O. DIVISION OF MATHEMATICS
AND STATISTICS
MELBOURNE, VICTORIA
AUSTRALIA

MARGARET WU
DEPARTMENT OF MATHEMATICS
MONASH UNIVERSITY
CLAYTON, VICTORIA
AUSTRALIA

R. G. STAUDTE, JR.
DEPARTMENT OF MATHEMATICS
LA TROBE UNIVERSITY
BUNDOORA, VICTORIA
AUSTRALIA 3083