The Annals of Statistics
2002, Vol. 30, No. 6, 1737–1759

THE INFLUENCE FUNCTION AND MAXIMUM BIAS OF TUKEY'S MEDIAN

BY ZHIQIANG CHEN[1] AND DAVID E. TYLER[2]

William Paterson University and Rutgers University

Tukey's median is among one of the earliest known high breakdown point multivariate location statistics. Aside from its breakdown point, though, little else appears to be known about its robustness properties. In this paper we investigate other aspects of Tukey's median, and in particular we derive and study its influence function and its maximum contamination bias function. When judged by these other robustness criteria, Tukey's median again proves to be highly robust.

1. Introduction and summary. In 1975, Tukey introduced the concept of data depth. This concept measures how deep points are embedded within a multivariate data set. It gives a more refined ordering of multivariate data than that obtained from convex peeling. Although other measures of data depth have subsequently been proposed, Tukey's depth still appears to be the most viable notion of depth; see, for example, the recent article by Zuo and Serfling (2000). The deepest or innermost point with respect to Tukey's depth is commonly viewed as a multivariate median, and so we refer to this deepest point as Tukey's median. Numerous papers have appeared which study various properties of Tukey's depth and median. Most notably, Donoho (1982) has shown that the breakdown point of Tukey's median can be as high as $\frac{1}{3}$, regardless of the dimension d of the data. In contrast, the breakdown point of the innermost point with respect to convex peeling is at best $\frac{1}{d+1}$. Aside from its breakdown point, little else appears to be known regarding the robustness properties of Tukey's median. In this paper, we consider other important robustness measures. In particular, we investigate the influence function and the maximum contamination bias function of Tukey's median. With respect to these criteria, Tukey's median again proves to be highly robust.

Our notation and definitions are formally set down in Section 2. The main results of the paper are given in Sections 3 and 4. Section 3 deals with continuous symmetric distributions, or more generally, the continuous halfspace symmetric distributions as defined in Zou and Serfling (2000). For such distributions, we show the following.

Received August 1999; revised August 2001.

[1]Supported in part by the Center for Research in the College of Science and Health, William Paterson University.

[2]Supported in part by NSF Grant IRI-99-87695.

AMS 2000 subject classifications. Primary 62G05, 62G20; secondary 62H12, 62G09, 62G10.

Key words and phrases. Contamination bias, contamination sensitivity, gross error sensitivity, halfspace depth, halfspace median, influence function, multivariate median, robustness, Tukey's depth.

1. Tukey's median has an influence function, which like the univariate median, is bounded and is constant along rays originating from the center of the distribution.
2. The gross error sensitivity of Tukey's median is never greater than the supremum of the gross error sensitivities of the univariate medians taken over all univariate projections of the distribution.
3. The maximum bias of Tukey's median under ε-contamination has a relatively simple form, namely the greatest distance between the deepest point and the convex contour corresponding to all points of depth $\frac{1-3\varepsilon}{2(1+\varepsilon)}$. This is particularly notable since maximum bias functions are usually difficult to derive even at the standard normal distribution, and seldom have simple forms.
4. The contamination sensitivity of Tukey's median is at least twice its gross error sensitivity, but it is never greater than twice the supremum of the contamination sensitivities of the univariate medians taken over all univariate projections of the distribution.

Section 4 considers absolutely continuous distributions in general. A simple upper bound is obtained for the maximum bias of Tukey's median under ε-contamination, and the contamination sensitivity is shown to be finite. A modest new result on the breakdown point of Tukey's median is also given.

Section 5 discusses some specific examples. In Section 5.1, we observe that for spherically symmetric distributions the gross error sensitivity and the maximum contamination bias of Tukey's median do not depend on the dimension d of the distribution. This contrasts with some other multivariate location statistics such as the M- and S-estimates. For these statistics, the gross error sensitivities tend to be of order $\sqrt{d}$. The gross error sensitivity and the maximum bias function of Tukey's median, though, can be dimension dependent for nonspherically symmetric distributions. This occurs, for example, for the multivariate distribution possessing i.i.d. Cauchy marginals; see Section 5.2. Finally, in Section 5.3, we illustrate our results on the maximum contamination bias function for nonsymmetric distributions by applying them to the uniform distribution having support within an equilateral triangle.

Technical proofs are reserved for the Appendix. Some modest new results on the uniqueness of the deepest point which arise as a byproduct of the proofs are given in Theorem 5.1.

2. Tukey's median. Let $X^{(n)} = \{x_1, x_2, \ldots, x_n\}$ represent a data set in $\Re^d$ with $d \geq 2$, and let $S^{d-1} = \{u \in \Re^d : u'u = 1\}$ be the set of all unit vectors in $\Re^d$. For a given point $x \in \Re^d$, consider the position of its univariate projection $u'x$ within the univariate data set $u'X^{(n)} = \{u'x_1, \ldots, u'x_n\}$ for some $u \in S^{d-1}$. Define the depth of $u'x$ within $u'X^{(n)}$ as the lesser of the number of points in $u'X^{(n)}$ which are no greater than $u'x$ and the number of points which are no less than $u'x$. The depth of the multivariate point $x \in \Re^d$ within $X^{(n)}$, as defined in Tukey (1975), is then the smallest such depth possible over all univariate projections of the data. More formally, Tukey's depth for a point x with respect to a data set $X^{(n)}$ is defined as

$$\text{depth}(x; X^{(n)}) = \min_{u \in S^{d-1}} nF_{u,n}(u'x), \tag{1}$$

where $F_{u,n}$ is the empirical cumulative distribution function of $X^{(n)}$. For a given $u \in S^{d-1}$, $nF_{u,n}(u'x)$ corresponds the number of points in $X^{(n)}$ which are no greater than $u'x$, whereas the number which are no less than $u'x$ is given by $nF_{-u,n}(-u'x)$. In one dimension, the point of maximal depth coincides with the univariate median, and so in higher dimensions

$$\arg\max_{x \in \Re^d} \text{depth}(x; X^{(n)}) \tag{2}$$

is generally referred to as a multivariate median, or more specifically as Tukey's median. Other generalizations of the median to higher dimensional spaces exist; see, for example, Small (1990) for a general review, as well as Zuo and Serfling (2000).

For our purposes, we need a functional or "population" version of Tukey's depth. Let X be a random vector with probability measure P on $\Re^d$, and let $F_u(y)$ represent the cumulative distribution function of $u'X$ for $u \in S^{d-1}$. A generalization of Tukey's depth to probability measures, up to a scaling factor, is given by

$$\pi(x; P) = \inf_{u \in S^{d-1}} F_u(u'x). \tag{3}$$

If P_n represents the empirical probability measure derived from $X^{(n)}$, then $\text{depth}(x, X^{(n)}) = n\pi(x; P_n)$.

Tukey's depth is sometimes referred to as the halfspace depth since it can be expressed as

$$\pi(x; P) = \inf_{u \in S^{d-1}} P\{H(u, x)\}, \tag{4}$$

where $H(u, x) = \{y \in \Re^d : u'y \leq u'x\}$ is the closed halfspace of $\Re^d$ whose boundary is orthogonal to $u \in S^{d-1}$ and passes through x. We utilize this representation of Tukey's depth for the proofs given in the Appendix.

A functional or "population" version of Tukey's median can be defined as

$$\arg\max_{x \in \Re^d} \pi(x; P), \tag{5}$$

which in general may not be unique. However, for $0 < \alpha < \Pi(P)$, where

$$\Pi(P) \equiv \max_{x \in \Re^d} \pi(x; P), \tag{6}$$

the "level" sets

$$L(\alpha; P) = \{x \in \Re^d : \pi(x; P) \geq \Pi(P) - \alpha\} \tag{7}$$

are nonempty convex compact sets and increasing in α. Hence, for any P the set

$$\text{Med}(P) \equiv \{x \in \Re^d : \pi(x; P) = \Pi(P)\} = \bigcap_{0<\varepsilon<\Pi(P)} L(\varepsilon; P) \tag{8}$$

is a nonempty convex compact set; see, for example, Small (1987) or Chen (1995). For the sample Tukey's median, Donoho and Gasko (1992) suggest taking an affine equivariant preserving average over $\text{Med}(P_n)$. This can also be done for general P by taking the expected value over the uniform distribution on $\text{Med}(P)$. That is, we can uniquely define Tukey's median (or the halfspace median) for a probability measure P on $\Re^d$ as

$$T(P) = \frac{\int_{\text{Med}(P)} y\, dy}{\int_{\text{Med}(P)} dy}, \tag{9}$$

where the integral is with respect to Lebesgue measure on the plane generated by $\text{Med}(P)$. The functional $T(P)$ is then affine equivariant. That is, if P^* represents the probability measure of $X^* = AX + b$ with A being nonsingular, then $T(P^*) = AT(P) + b$. There are other choices for defining an affine equivariant preserving average over $\text{Med}(P)$. The results of this paper hold for any such choice.

3. Symmetric distributions.

3.1. *Halfspace symmetry.* Hereafter, we assume that P is absolutely continuous with respect to Lebesgue measure on $\Re^d$. In order to derive the influence function of T at P, we will need $\arg\max_{x\in\Re^d} \pi(x; P)$ to be unique. This occurs, for example, if P has a probability distribution which is symmetric about some point $\mu \in \Re^d$; that is, if the distribution of $(X - \mu)$ and $-(X - \mu)$ are the same. A symmetric distribution about μ is sometimes referred to as being centrosymmetric about μ; see, for example, Donoho and Gasko (1992). In this section we assume a more general notion of symmetry for P, namely we assume only that P is halfspace symmetric, which as defined in Zuo and Serfling (2000), means there exists a $\mu \in \Re^d$ for which any plane passing through μ divides $\Re^d$ into halfspaces of equal probability. More formally, it means that there exists a $\mu \in \Re^d$ such that $P\{H(u, \mu)\} = P\{H(-u, \mu)\}$ for all $u \in S^{d-1}$. It readily follows that the class of halfspace symmetric distributions contains the class of symmetric distributions. Furthermore, for halfspace symmetric distributions, $\arg\max_{x\in\Re^d} \pi(x; P)$ is unique and is equal to μ, $T(P) = \mu$ and $\Pi(P) = \frac{1}{2}$. These last statements can be found in either Small (1987), Chen (1995) or Zuo and Serfling (2000).

Since $T(P)$ is affine equivariant, unless stated otherwise, we assume without loss of generality that $T(P) = 0$ and hence the "center" $\mu = 0$ for a halfspace symmetric distribution. Zuo and Serfling (2000) note when P is continuous, as is assumed here, then halfspace symmetry is equivalent to angular symmetry, but not necessarily otherwise. This implies that for $\mu = 0$ the directional vector $\frac{X}{|X|}$ has a symmetric distribution about the origin, where here and throughout, $|\cdot|$ refers to the Euclidean norm in $\Re^d$. Angular symmetry is at times also referred to as directional symmetry; see, for example, Randles (2000). The terminology directional symmetry is perhaps more descriptive since it reflects that it is the directional vector $\frac{X}{|X|}$, rather than any angle created by the directional vector, which has a symmetric distribution about the origin. Alternatively, using standard terminology from the directional data literature, we can say that the directional vector $\frac{X}{|X|}$ has an antipodal symmetric distribution; see, for example, Mardia (1972) or Watson (1983).

3.2. *The influence function and gross error sensitivity.* To understand the local stability of a statistic, one often studies its influence function. Recall that the influence function of T at P as a function of $x \in \Re^d$, as defined, for example, in Hampel, Ronchetti, Rousseeuw and Stahel (1986), is given by

$$IF(x; T, P) = \lim_{\varepsilon \to 0} \frac{T(P(\varepsilon, x)) - T(P)}{\varepsilon}, \tag{10}$$

where δ_x is the point mass probability measure at $x \in \Re^d$, and $P(\varepsilon, x) = (1-\varepsilon)P + \varepsilon\delta_x$ for $\varepsilon \in [0, 1]$. In order to obtain results on the influence function, we first establish some results for $T(P(\varepsilon, x))$. For $0 \le \alpha < \frac{1}{2}$ and $u \in S^{d-1}$, let $r(\alpha, u)$ be defined as the radius of the level set $L(\alpha; P)$ along the direction u, or more specifically, the length of the line segment $L(\alpha; P) \cap \{tu : t \ge 0\}$. It is understood that $r(\alpha, u)$ depends upon P, which for convenience we do not include in the notation. The influence function is obtained as a direct consequence of the following theorem. This key theorem is also applied later in deriving the maximum contamination bias function of Tukey's median.

THEOREM 3.1. *Suppose P is absolutely continuous with respect to Lebesgue measure on $\Re^d$ and is halfspace symmetric about the origin. For an arbitrary fixed $x \in \Re^d$ and $0 \le \varepsilon < \frac{1}{3}$,*

$$T(P(\varepsilon, x)) = \begin{cases} x, & \text{for } x \text{ in the interior of } L\left(\dfrac{\varepsilon}{1-\varepsilon}; P\right), \\ \dfrac{1}{2} r\left(\dfrac{\varepsilon}{1-\varepsilon}, \dfrac{x}{|x|}\right)\dfrac{x}{|x|}, & \text{for } x \notin L\left(\dfrac{\varepsilon}{1-\varepsilon}; P\right) \text{ or on its boundary.} \end{cases}$$

To obtain the influence function of T at P, we note that for any fixed $x \neq 0$, there is a small enough ε such that $x \notin L(\frac{\varepsilon}{1-\varepsilon}; P)$.

COROLLARY 3.1. *Under the conditions of Theorem 3.1, for $x \neq 0$*

$$IF(x; T, P) = \frac{1}{2}\lambda\left(\frac{x}{|x|}\right)\frac{x}{|x|}$$

provided $\lambda(\frac{x}{|x|}) = \lim_{\alpha \to 0} r(\alpha, \frac{x}{|x|})/\alpha$ *exists. For* $x = 0$, $IF(x; T, P) = 0$.

From Theorem 3.1, we observe that for a fixed ε, $T(P(\varepsilon, x))$ does not increase montonically along rays originating from the origin. Rather it first increases linearly along a ray and then redescends to a fixed value. This curious redescending property of Tukey's median is not apparent from the influence function alone, which is constant along rays originating from the origin.

To help illustrate the above results, consider the special case when P has an elliptically symmetric distribution, which without loss of generality we take to be symmetric about the origin. That is, suppose P has a density of the form

$$f(x; \Sigma) = |\Sigma|^{-1/2} g(x'\Sigma^{-1}x) \tag{11}$$

for some function $g \geq 0$, and with Σ being a symmetric positive definite matrix of order d. When $\Sigma = I_d$, which corresponds to P having a spherically symmetric distribution, the distribution of the univariate projection $Y = u'X$ is the same for any $u \in S^{d-1}$, and so for this case we can denote the cumulative distribution function of Y simply by $F_o(y)$ and the density function of Y by $f_o(y)$. For general Σ, the cumulative distribution function $F_u(y)$ and density $f_u(y)$ of $Y = u'X$ can be expressed in terms of F_o and f_o respectively. That is,

$$F_u(x) = F_o\big(x/\sqrt{u'\Sigma u}\big) \quad \text{and} \quad f_u(x) = \frac{f_o(x/\sqrt{u'\Sigma u})}{\sqrt{u'\Sigma u}}. \tag{12}$$

The α-level sets $L(\alpha; P)$ for elliptical distributions are given by the interiors and boundaries of ellipses, specifically

$$L(\alpha; P) = \{x : x'\Sigma^{-1}x \leq r_o^2(\alpha)\}, \tag{13}$$

where $r_o(\alpha) = F_o^{-1}(\frac{1}{2} + \alpha)$. This then gives $r(\alpha, u) = r_o(\alpha)/\sqrt{u'\Sigma^{-1}u}$, and so we have, for $u \in S^{d-1}$,

$$\lambda(u) = \{f_o(0)\sqrt{u'\Sigma^{-1}u}\}^{-1}. \tag{14}$$

For spherically symmetric distributions with $\Sigma = I_d$, $\lambda(u) = 1/f_o(0)$. For the standard normal distributions $\lambda(u) = \sqrt{2\pi}$. In general, the density function $f_o(y)$ does not have a simple closed form expression in terms of the function g in (11). For the important case $f_o(0)$, we have the relationship

$$f_o(0) = \omega_{d-1} \int_0^\infty g(r^2) r^{d-2}\, dr > 0, \tag{15}$$

where $\omega_{d-1} = 2\pi^{(d-1)/2}\Gamma(\frac{d-2}{2})$. Proofs of the above statements concerning the elliptical distributions are given in the Appendix.

When P has an elliptically symmetric distribution, it follows from (14) and (15) that the influence function is bounded. Unlike the univariate case, this holds even if the density of P vanishes at the origin since $f_o(0) > 0$ always holds even if $g(0) = 0$. The gross error sensitivity of T at P is given by

$$\gamma(T; P) = \sup_{x \in \Re^d} |IF(x; T, P)| = \frac{1}{2} \sup_{u \in S^{d-1}} \lambda(u) = \frac{\sqrt{\lambda_1(\Sigma)}}{2 f_o(0)}, \tag{16}$$

where $\lambda_1(\Sigma)$ represents the largest eigenvalue of Σ. The most influential points, that is, points x which maximize the Euclidean norm of the influence function, are those which are proportional to the eigenvectors of Σ corresponding to the largest eigenvalue $\lambda_1(\Sigma)$. This last statement follows immediately from application of the Cauchy–Schwarz inequality. When $\Sigma \neq I_d$, it is customary to use the self-standardizing gross error sensitivity based on the norm $|x|_\Sigma = \sqrt{x' \Sigma^{-1} x}$ rather than use the Euclidean norm; see, for example, Hampel, Ronchetti, Rousseeuw and Stahel (1986). This then gives

$$\gamma(T; P, \Sigma) = \sup_{x \in \Re^d} |IF(x; T, P)|_\Sigma = 1/\{2 f_o(0)\}. \tag{17}$$

Note that this is equivalent to first transforming P so that it has a spherically symmetric distribution and then evaluating the gross error sensitivity using the Euclidean norm. When using the self-standardizing gross error sensitivity at an elliptically symmetric distribution, all points $x \neq 0$ are considered equally influential.

In general, for any halfspace symmetric distribution, we show in the Appendix that

$$r(\alpha, u) \leq F_u^{-1}\left(\tfrac{1}{2} + \alpha\right) \tag{18}$$

and so, since $F_u^{-1}(1/2) = 0$, we have

$$\lambda(u) \leq \lim_{\alpha \to 0} \frac{F_u^{-1}(\frac{1}{2} + \alpha)}{\alpha} = \{f_u(0)\}^{-1}, \tag{19}$$

where $f_u(y)$ represents the density of the univariate projection $Y = u'X$ for $u \in S^{d-1}$, or to be more precise, it represents the derivative from above of $F_u(y)$. As a side note, it is interesting to observe that for elliptically symmetric distributions, equality holds in (18) and (19) if and only if u is an eigenvector of Σ. [This can be verified by using the special expressions for F_u, f_u, $r(\alpha, u)$ and $\lambda(u)$ given in (12), (13) and (14) whenever P has an elliptically symmetric distribution, and then using the Cauchy–Schwarz inequality to note that $u'\Sigma^{-1}u \, u'\Sigma u \geq 1$ for $u \in S^{d-1}$, with equality if and only if u is an eigenvector of Σ.] For any halfspace symmetric distribution, inequality (19) can be used to obtain the following upper bound on the gross error sensitivity of Tukey's median.

THEOREM 3.2. *Under the conditions of Corollary* 3.1,

$$\gamma(T; P) = \sup_x |IF(x, T, P)| = \tfrac{1}{2} \sup_{u \in S^{d-1}} \lambda(u) \leq \left\{ 2 \inf_{u \in S^{d-1}} f_u(0) \right\}^{-1} < \infty,$$

provided $\inf_{u\in S^{d-1}} f_u(0) > 0$.

The condition $\inf_{u\in S^{d-1}} f_u(0) > 0$ is considerably weaker than assuming the density of P is positive at the origin. Due to the compactness of S^{d-1}, if we make the additional assumption that $f_u(0)$ is a continuous function of $u \in S^{d-1}$, then the condition holds if and only if $f_u(0) > 0$ for all $u \in S^{d-1}$. That is, it holds provided no univariate marginal density vanishes at the origin, and does not hold otherwise. This means that rather than demanding that the density of P be nonzero at the origin, we demand only that the density of P not be almost surely zero on any $d-1$ dimensional subspace of $\Re^d$.

In one dimension, the gross error sensitivity of the median is well known to be equal to the inverse of twice the univariate density evaluated at zero; see, for example, Hampel, Ronchetti, Rousseeuw and Stahel (1986). Suppose we let $\mathrm{med}_u(P)$ represent the functional corresponding to the median of the univariate projection $u'X$. The gross error sensitivity of $\mathrm{med}_u(P)$ is then given by $\gamma(\mathrm{med}_u; P) = \{2f_u(0)\}^{-1}$ and so we note that the inequality in Theorem 3.2 states

$$\gamma(T; P) \le \sup_{u\in S^{d-1}} \gamma(\mathrm{med}_u; P); \tag{20}$$

that is, the gross error sensitivity of Tukey's median is never worse than the supremum of the gross error sensitivities of the univariate median over all possible univariate projections of the distribution. We have equality in (20) for elliptically symmetric distributions.

The influence function of Tukey's median at P does not exist if $\lim_{\alpha\to 0} r(\alpha, \frac{x}{|x|})/\alpha$ does not exist. However, from (18) we see that $r(\alpha, \frac{x}{|x|})/\alpha$ is bounded above by a convergent sequence and so for $u \in S^{d-1}$,

$$\lambda^*(u) = \limsup_{\alpha\to 0} r(\alpha, u)/\alpha \tag{21}$$

exists. If we then apply this result to the more general definition of gross error sensitivity given, for example, in Hampel, Ronchetti, Rousseeuw and Stahel (1986), namely

$$\gamma(T; P) \equiv \sup_{x\in\Re^d} \limsup_{\varepsilon\to 0} \frac{|T(P(\varepsilon, x)) - T(P)|}{\varepsilon}, \tag{22}$$

then we have the following more general version of Theorem 3.2.

THEOREM 3.3. *Under the conditions of Theorem* 3.1,

$$\gamma(T; P) = \tfrac{1}{2} \sup_{u\in S^{d-1}} \lambda^*(u) \le \left\{2 \inf_{u\in S^{d-1}} f_u(0)\right\}^{-1} < \infty$$

provided $\inf_{u\in S^{d-1}} f_u(0) > 0$.

A self-standardizing gross error sensitivity can be defined in general by $\gamma(T; P, \Sigma(P)) = \sup_x |IF(x, T, P)|_{\Sigma(P)}$, where $\Sigma(P)$ is some well defined affine equivariant scatter functional. That is, $\Sigma(P)$ represents some symmetric positive definite matrix functional of order d possessing the property $\Sigma(P^*) = A\Sigma(P)A'$, where A is nonsingular and P^* represents the probability measure of $X^* = AX + b$. The self-standardizing gross error sensitivity is affine invariant in the sense that $\gamma(T; P, \Sigma(P)) = \gamma(T; P^*, \Sigma(P^*))$, and so it is equivalent to using $\gamma(T; P)$ after the distribution P is standardized by an affine transformation so that $\Sigma(P) = I_d$.

3.3. *Contamination bias and sensitivity.* The concept of gross error sensitivity measures the maximum effect that an infinitesimal amount of point-mass contamination can have on a functional. A stronger robustness concept is to measure the maximum effect or bias that any type of contamination can have on a functional. The maximum contamination bias function for $T(P)$ is defined to be

$$B(\varepsilon; T, P) \equiv \sup_Q |T(P(\varepsilon, Q)) - T(P)| \tag{23}$$

where $P(\varepsilon, Q) = (1 - \varepsilon)P + \varepsilon Q$ and Q is any arbitrary probability measure on $\Re^d$; see, for example, He and Simpson (1993). As with the gross error sensitivity, an affine invariant version of the maximum bias function can be defined by replacing the Euclidean norm $|\cdot|$ in (23) by $|\cdot|_{\Sigma(P)}$. This is equivalent to applying (23) after standardizing the distribution so that $\Sigma(P) = I_d$.

The maximum bias function is related to the breakdown point, which is a measure of global robustness, as well as to the contamination sensitivity, which is a measure of local robustness. The breakdown point of T at P over contamination neighborhoods is defined as

$$\varepsilon^*(T; P) \equiv \inf\{\varepsilon : B(\varepsilon; T, P) = \infty\}, \tag{24}$$

and the contamination sensitivity is defined as

$$\gamma^*(T; P) \equiv \limsup_{\varepsilon \to 0} B(\varepsilon; T, P)/\varepsilon. \tag{25}$$

Under certain regularity conditions, the contamination sensitivity and the gross error sensitivity are equal; see Hampel, Ronchetti, Rousseeuw and Stahel (1986) for further discussion. In general, though, it readily follows that

$$\gamma^*(T; P) \ge \sup_{x \in \Re^d} \left\{ \limsup_{\varepsilon \to 0} |T(P(\varepsilon, x) - T(P)|/\varepsilon \right\} = \gamma(T; P). \tag{26}$$

As we will see, for Tukey's median this inequality is strict.

In many problems, the contaminating distribution which produces the maximal bias in (23) is often a point mass distribution; see, for example, Martin, Yohai and Zamar (1989). As the following theorem states, this is also true for Tukey's median at halfspace symmetric distributions. Here, for a set $A \subset \Re^d$, we set $\|A\| \equiv \sup_{y \in A} |y|$.

THEOREM 3.4. *Under the conditions of Theorem* 3.1 *and for* $\varepsilon < \frac{1}{3}$,

$$B(\varepsilon; T, P) = \sup_x |T(P(\varepsilon, x)) - T(P)| = \left\| L\left(\frac{\varepsilon}{1-\varepsilon}; P\right) \right\|.$$

Furthermore, $\varepsilon^*(T; P) = \frac{1}{3}$.

Note that $\|L(\alpha; P)\| = \sup_{u \in S^{d-1}} r(\alpha, u)$ and hence under the conditions of Theorem 3.1 we have

$$\gamma^*(T; P) \geq \sup_{u \in S^{d-1}} \limsup_{\alpha \to 0} r(\alpha, u)/\alpha \geq 2\gamma(T; P); \tag{27}$$

that is, the contamination sensitivity differs from the gross error sensitivity by at least a factor of 2. This is attributable to the different orders in which the supremums and the limits are taken in their respective definitions. In particular, the first part of the expression for $T(P(\varepsilon, x))$ in Theorem 3.1 applies when deriving the contamination sensitivity, while the second part applies when deriving the gross error sensitivity.

For the special case of elliptically symmetric distributions, the special form of $r(\alpha, u)$ given after (13) can be used to obtain relatively simple expressions for the maximum contamination bias function and the contamination sensitivity. These are, respectively,

$$B(\varepsilon; T, P) = \sqrt{\lambda_1(\Sigma)} F_o^{-1}\left\{\frac{1+\varepsilon}{2(1-\varepsilon)}\right\} \quad \text{and} \quad \gamma^*(T; P) = \frac{\sqrt{\lambda_1(\Sigma)}}{f_o(0)}. \tag{28}$$

Note that in this case we obtain equality in (27).

In general, we can bound the contamination sensitivity by applying the bound (18) and by assuring that the lim sup and the sup can be interchanged in (25) and (27). This is true if $\lim_{\alpha \to 0} F_u^{-1}(\frac{1}{2} + \alpha)/\alpha \to 1/f_u(0)$ uniformly in $u \in S^{d-1}$, which in turn holds if $f_u(0)$ is continuous in $u \in S^{d-1}$. We thus have the following theorem.

THEOREM 3.5. *Under the conditions of Theorem* 3.3,

$$\gamma^*(T, P) \leq \left\{ \inf_{u \in S^{d-1}} f_u(0) \right\}^{-1} < \infty.$$

provided $f_u(0)$ *is continuous in* $u \in S^{d-1}$.

4. General distributions. In this section we drop the symmetry assumption and show that the contamination sensitivity of Tukey's median is finite in general. We still assume that P is absolutely continuous with respect to Lebesgue measure on $\Re^d$. We begin by giving a general bound on the maximum bias function, along with some modest new results on the breakdown point of Tukey's median. Note that we still assume without loss of generality that $T(P) = 0$. Also, recall that $\Pi(P) = \max_{x \in \Re^d} \pi(x; P)$.

THEOREM 4.1. *If P is absolutely continuous with respect to Lebesgue measure on $\Re^d$, then for $\varepsilon < \frac{\Pi(P)}{1+\Pi(P)}$,*

$$B(\varepsilon; T, P) \le \left\| L\left(\frac{\varepsilon}{1-\varepsilon}; P\right)\right\|.$$

Furthermore, $\frac{\Pi(P)}{1+\Pi(P)} \le \varepsilon^*(T; P) \le \frac{1}{3}$.

Donoho and Gasko (1992) give a finite sample version of the breakdown point of Tukey's median and show that its limiting value is $\frac{1}{3}$ for symmetric distributions and between $\frac{1}{d+1}$ and $\frac{1}{3}$ for distributions in general. They also show that $\Pi(P) \ge \frac{1}{d+1}$, which implies $\frac{\Pi(P)}{1+\Pi(P)} \ge \frac{1}{d+2}$. Although Theorem 4.1 is sufficient for showing that the contamination sensitivity is bounded, the bound on the maximum bias function can be tightened so that we obtain $\varepsilon^*(T; P) \ge \frac{1}{d+1}$. To do so, let $\Pi(\varepsilon; P) \equiv \inf_Q \Pi(P(\varepsilon, Q))$, which is a continuous nonincreasing function of ε, and define $\varepsilon_{\mathrm{L}}(T; P)$ to be the solution to the equation $\varepsilon = \Pi(\varepsilon; P)$.

THEOREM 4.2. *If P is absolutely continuous with respect to Lebesgue measure on $\Re^d$, then for $\varepsilon < \varepsilon_L(T; P)$,*

$$B(\varepsilon; T, P) \le \left\| L\left(\frac{\varepsilon}{1-\varepsilon} - \delta(\varepsilon; P); P\right)\right\|,$$

where $\delta(\varepsilon; P) = \frac{\Pi(\varepsilon; P) - (1-\varepsilon)\Pi(P)}{1-\varepsilon} \ge 0$. Furthermore, $\varepsilon_L(T; P) \ge \frac{\Pi(P)}{1+\Pi(P)}$ and $\frac{1}{d+1} \le \varepsilon_L(T; P) \le \varepsilon^*(T; P) \le \frac{1}{3}$.

The form of $\delta(\varepsilon; P)$ is somewhat complex. A simple lower bound for it can be obtained by recalling $\Pi(\varepsilon; P) \ge \frac{1}{d+1}$, and hence $\delta(\varepsilon; P) \ge \delta_o(\varepsilon; P) \equiv \max\{0, \frac{1}{(d+1)(1-\varepsilon)} - \Pi(P)\}$. This lower bound can then be used in Theorem 4.1 in place of $\delta(\varepsilon; P)$ to obtain a simpler but cruder upper bound for the maximum bias function. An example illustrating this is given in Section 5.3. Note that for $d = 2$ we always have $\varepsilon^*(T; P) = \frac{1}{3}$.

Finally, in order to show that the contamination sensitivity is bounded, it is not necessary for $\arg\max_{x \in \Re^d} \pi(x; P)$ to be unique. However, we still need

$\inf_{u \in S^{d-1}} f_u(0) > 0$. The proof of this seemingly simple result, which is given in the Appendix, is rather challenging.

THEOREM 4.3. *If P is absolutely continuous with respect to Lebesgue measure on $\Re^d$ and $\inf_{u \in S^{d-1}} f_u(0) > 0$, then $\gamma^*(T; P) < \infty$.*

5. Some examples with discussion.

5.1. *Spherically symmetric distributions.* We have already noted the relatively simple form of the influence function and the maximum contamination bias function of Tukey's median at elliptically symmetric distributions. We now examine these functions in more detail. For simplicity, we restrict our attention to the subclass of spherically symmetric distributions. That is, unless stated otherwise we assume in this subsection that P has a density which is given by (11) with $\Sigma = I_d$. Recall that if X has a spherically symmetric distribution, then the distribution of the univariate projection $u'X$ is the same for any $u \in S^{d-1}$. For any $u \in S^{d-1}$, the cumulative distribution function and the density function of $u'X$ are again denoted by F_o and f_o respectively.

To begin, we note that $T_u(P) = u'T(P)$ represents a measure of univariate location for $u'X$, and so it is natural to compare it to the univariate median of $u'X$. Without loss of generality, consider simply $T_1(P)$ versus $\text{med}_1(P) = \text{median}(X_1)$. Their influence functions at P as a function of $x \in \Re^d$ are, respectively,

$$\begin{aligned} IF(x; T_1, P) &= \frac{1}{2 f_o(0)} \frac{x_1}{|x|} \quad \text{and} \\ IF(x; \text{med}_1, P) &= \frac{1}{2 f_o(0)} \operatorname{sign}(x_1). \end{aligned} \tag{29}$$

Although the gross error sensitivities are the same, namely $\gamma(T_1; P) = \gamma(\text{med}_1; P) = \{2 f_o(0)\}^{-1}$, the influence function for T_1 is always smaller then that of med_1 whenever x does not lie along the first coordinate. The functional T_1 makes use of the information in all the coordinates in assessing the "outlyingness" of x, whereas med_1 uses only the information in x_1. The univariate median, though, always fares better when judged in terms of the maximum bias function since

$$B(\varepsilon; T_1, P) = F_o^{-1}\left\{\frac{1+\varepsilon}{2(1-\varepsilon)}\right\} > B(\varepsilon; \text{med}_1, P) = F_o^{-1}\left\{\frac{1}{2(1-\varepsilon)}\right\} \tag{30}$$

for any $\varepsilon > 0$. The expression for $B(\varepsilon; \text{med}_1, P)$ is well known and can be found, for example, in Huber (1981). For the contamination sensitivities, we likewise have $\gamma^*(T_1; P) = \{f_o(0)\}^{-1} > \gamma^*(\text{med}_1; P) = \{2 f_o(0)\}^{-1}$. If we consider the bias itself under a point mass contamination at x, then the bias of T_1 can still be considerably less than the bias of the marginal median med_1 whenever $x \in \Re^d$

does not lie along the first coordinate. Specifically, under an ε contamination at x, T_1 and med_1 change respectively from zero to

$$T_1(P(\varepsilon, x)) = \begin{cases} x_1, & |x| < r_\varepsilon, \\ \dfrac{1}{2}\dfrac{x_1}{|x|} r_\varepsilon, & |x| \ge r_\varepsilon, \end{cases}$$

and

$$\mathrm{med}_1(P(\varepsilon, x)) = \begin{cases} x_1, & |x_1| < b_\varepsilon, \\ \mathrm{sign}(x_1) b_\varepsilon, & |x_1| \ge b_\varepsilon, \end{cases}$$

where $r_\varepsilon = B(\varepsilon; T_1, P)$ and $b_\varepsilon = B(\varepsilon; \mathrm{med}_1, P)$. For $|x| \ge r_\varepsilon$ and $|x_1| \ge b_\varepsilon$, the term $x_1/|x|$ can be arbitrarily small.

A curious property of Tukey's median is that its gross error sensitivity, maximum contamination bias function and contamination sensitivity, which are, respectively,

$$\gamma(T; P) = \frac{1}{2 f_o(0)}, \qquad B(\varepsilon; T, P) = F_o^{-1}\left\{\frac{1+\varepsilon}{2(1-\varepsilon)}\right\}$$

and

$$\gamma^*(T; P) = \frac{1}{f_o(0)}$$

are the same as those for T_1 itself, or more generally as those of T_u for any fixed $u \in S^{d-1}$. Moreover, these expressions remain the same as the dimension d increases as long as the marginal distributions remain fixed. This property is unusual for multivariate location functionals. For example, the gross error sensitivity, the maximum bias function and the contamination sensitivity of the joint marginal medians (which is not affine equivariant), that is, of $T_M(P) = (\mathrm{med}_1(P), \dots, \mathrm{med}_d(P))'$, increase at the rate of $\sqrt{d}$. Specifically,

$$\gamma(T_M; P) = \frac{\sqrt{d}}{2 f_o(0)}, \qquad B(\varepsilon; T_M, P) = \sqrt{d} F_o^{-1}\left\{\frac{1}{2(1-\varepsilon)}\right\}$$

and

$$\gamma^*(T_M; P) = \frac{\sqrt{d}}{2 f_o(0)}.$$

The expressions above follow from considering point mass contaminating distributions at points which are proportional to the vector $(1, \dots, 1)'$. Note that the gross error sensitivity of Tukey's median is always smaller than that of the joint marginal medians for $d \ge 2$, and its contamination sensitivity is less than or equal to that of the joint marginal medians for $d \ge 4$. A comparison of the maximum contamination bias function of Tukey's median to that of the joint marginal medians for different dimensions d whenever $P = P_Z$, the standard multivariate normal

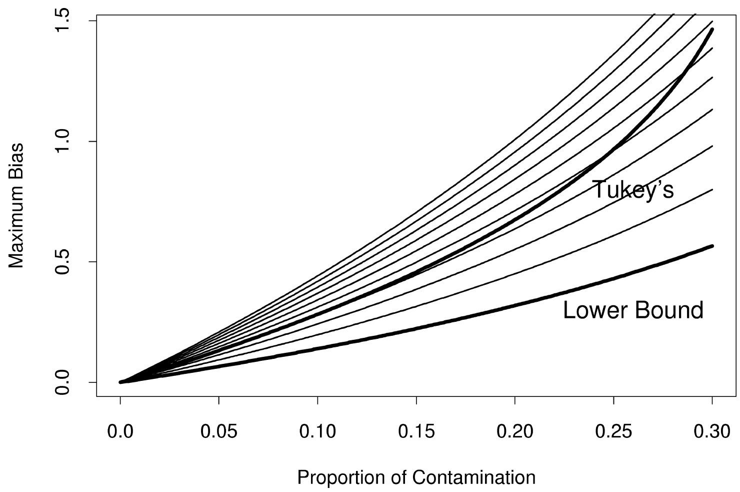

FIG. 1. *Plots of* ε *versus* $B(\varepsilon; T, P_Z)$ (*upper heavy line*), $B(\varepsilon; T_M, P_Z)$ *for* $d = 2, \ldots, 10$ (*light lines, in ascending order*) *and* $B_L(\varepsilon; P_Z)$ (*lower heavy line*).

distribution, is given in Figure 1. Although the breakdown point of Tukey's median is less than the breakdown point of the joint marginal medians, namely $1/3$ versus $1/2$, the maximum bias of Tukey's median at P_Z is smaller for $d > 4$ even for contamination proportions as high as $\varepsilon = 0.24$.

The lower bound indicated in Figure 1 follows from some general results for lower bounds on the maximum contamination bias of a functional given by He and Simpson (1993). These general results have recently been applied by Adrover and Yohai (2002) to obtain a lower bound for the maximum contamination bias, as well as a lower bound for the contamination sensitivity, of any affine equivariate location functionals at some unimodal spherically symmetric distribution. These lower bounds, which may not be strict, are respectively

$$B_L(\varepsilon; P) \equiv F_o^{-1}\left(\frac{1}{2(1-\varepsilon)}\right) \quad \text{and} \quad \gamma_L^*(P) \equiv \frac{1}{2 f_o(0)}. \tag{31}$$

Note that these bounds are equal to the maximum contamination bias function and the contamination sensitivity of the univariate median respectively. The contamination sensitivity of Tukey's median is only twice the corresponding lower bound, and so for smooth F_o, the maximum contamination bias of Tukey's median is approximately twice its corresponding lower bound for small values of ε. The ratios of the maximum bias of Tukey's median at P_Z to the lower bound for $\varepsilon = 0.05$, 0.10, 0.15, 0.20, 0.25 and 0.30 are respectively 2.004, 2.020, 2.053, 2.117, 2.246 and 2.590. The lower bound in Figure 1 also applies to the joint marginal medians since Croux, Haesbroeck and Rousseeuw (2001) have shown that the bound holds at P_Z for any translation equivariant multivariate location functional rather than for just affine equivariant functionals.

The dimension-free property of the gross error sensitivity, the maximum contamination bias and the contamination sensitivity of Tukey's median sets it apart not only from the joint marginal medians but also from other robust estimates of multivariate location such as the M-estimates, the S-estimates, the MM-estimates and the constrained M-estimates. All of these latter estimates have, for example, gross error sensitivities that tend to increase at the rate of $\sqrt{d}$; see Maronna (1976), Lopuhaä (1989), Lopuhaä (1992) and Kent and Tyler (1996), respectively. The following simple heuristic argument shows that this $\sqrt{d}$ rate holds, under certain regularity conditions, at the multivariate normal model for any asymptotic normal estimate of multivariate location. Suppose $T_o(F_n)$ is an estimate of μ based on a random sample of size n and that under the $\mathrm{Normal}_d(\mu, I_d)$ model

$$\sqrt{n}\{T_o(F_n) - \mu\} \to \mathrm{Normal}_d(0, \Gamma_o)$$

in distribution. Under broad regularity conditions,

$$\Gamma_o = E_P\big[IF(X; T_o, P)IF(X; T_o, P)'\big]$$

and the asymptotic variance covariance matrix of $T_o(F_n)$ cannot be less than that of the sample mean vector; that is, $\Gamma_o \geq I_d$ under the usual ordering of symmetric positive-definite matrices. Taking the trace of both sides of this inequality gives

$$\gamma(T_o, P) \geq E_P\big[\|IF(X; T_o, P)\|^2\big]^{1/2} \geq \sqrt{d}. \tag{32}$$

The above argument does not apply to Tukey's median since, as shown by Nolan (1999) and by Bai and He (1999), Tukey's median does not have an asymptotic normal distribution. This emphasizes a limitation in the definitions of the influence function and the maximum bias function. Namely, they are defined only with respect to contamination neighborhoods. Even though F_n may be "close" to F, it does not lie within some contamination neighborhood of F. In general, the influence function and the maximum bias function may not completely reflect the stability of a functional near a particular model. Nevertheless, under contamination, Tukey's median is remarkably stable.

5.2. *Independent Cauchy marginals.* In this subsection, we consider the random vector $X = (X_1, \dots, X_d)'$ having density

$$f(x_1, \dots, x_d) = \pi^{-d} \prod_{i=1}^{d} (1 + x_i^2)^{-1}. \tag{33}$$

That is, X is formed from independent and identically distributed standard Cauchy marginals. We denote the probability measure associated with X by P_C. Note that the distribution of X is symmetric about the origin, but it is not spherically symmetric. Although P_C and a spherically symmetric Cauchy distribution both have Cauchy marginals, the marginals are not independent for the latter. Also, although any linear combination of either distribution has a univariate Cauchy distribution, the spread of the univariate Cauchy distribution for $u'X$ is not the same for all $u \in S^{d-1}$ under P_C, whereas they are the same under a spherically symmetric Cauchy distribution.

Under P_C, it is well known that the distribution of $Y = u'X$ is the same as the distribution of $\sigma_u X_1$, where $\sigma_u = \|u\|_1$, and hence Y has density

$$f_u(y) = (\pi\sigma_u)^{-1}\left(1 + \frac{y^2}{\sigma_u^2}\right)^{-1}. \tag{34}$$

Here, for $a \in \Re^d$, $\|a\|_1 \equiv \sum_{i=1}^d |a_i|$ is the L_1 norm. In this subsection, we also use the L_2 norm and the L_∞ norm. To avoid confusion, we now denote the Euclidean or L_2 norm by $\|a\|_2$, and the L_∞ norm by $\|a\|_\infty \equiv \max_{i=1,\dots,d} |a_i|$.

In the bivariate case, Rousseeuw and Ruts [(1999), Proposition 13] derive the depth function for P_C. Their results can be extended to general d as follows. Since $P_C(H(u,x)) = P(u'X \le u'x) = P(X_1 \le \frac{u'x}{\|u\|_1})$, application of Holder's inequality implies $P_C(H(u,x))$ is minimized when $u'x = -\|u\|_1\ \|x\|_\infty$ and hence

$$\pi(x, P_C) = P(X_1 \le -\|x\|_\infty) = \frac{1}{2} - \frac{1}{\pi}\arctan(\|x\|_\infty). \tag{35}$$

From this it follows that the convex set corresponding to all points of depth greater than or equal to $\frac{1}{2} - \alpha$, that is, the level set $L(\alpha, P_\mathrm{C}) = \{x : \|x\|_\infty \le \tan(\pi\alpha)\}$, consists of those points on or inside a cube centered at the origin, with sides parallel to the coordinate axes and half-lengths equal to $\tan(\pi\alpha)$. Figure 2 shows these depth contours, along with the contours of the density (33), for the bivariate case.

For P_C, the general results on Tukey's median for symmetric distributions given in Section 2.2 can be expressed as follows:

$$r(\alpha, u) = \tan(\pi\alpha)/\|u\|_\infty, \tag{36}$$

$$T(P(\varepsilon, x)) = \begin{cases} x, & \text{for } \|x\|_\infty < \tan\left(\dfrac{\pi\varepsilon}{1-\varepsilon}\right), \\[2ex] \dfrac{1}{2}\tan\left(\dfrac{\pi\varepsilon}{1-\varepsilon}\right)\dfrac{x}{\|x\|_\infty}, & \text{for } \|x\|_\infty \ge \tan\left(\dfrac{\pi\varepsilon}{1-\varepsilon}\right) \end{cases} \tag{37}$$

and

$$IF(x, T, P_\mathrm{C}) = \frac{\pi}{2}\frac{x}{\|x\|_\infty}. \tag{38}$$

The L_2 norm of the influence function is maximized when $|x_1| = \cdots = |x_d|$, resulting in a gross error sensitivity of $\gamma(T, P_C) = \frac{\pi\sqrt{d}}{2}$. For the maximum bias

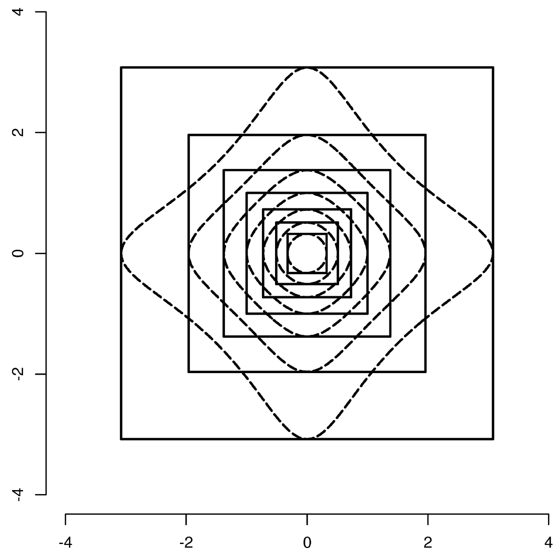

FIG. 2. *The depth contours* (*solid lines*) *and the density contours* (*dashed lines*) *of* P_C *for* $d = 2$. *The depth contours, from the outermost to the innermost square, represent depths of* 0.10, 0.15, 0.20, 0.25, 0.30, 0.35 *and* 0.40, *respectively*.

function of Tukey's median at P_C we have

$$B(\varepsilon, T, P_C) = \sup_{u \in S^{d-1}} r\left(\frac{\varepsilon}{1-\varepsilon}, u\right) = \sqrt{d}\tan\left(\pi\frac{\varepsilon}{1-\varepsilon}\right), \tag{39}$$

with the contamination sensitivity being $\gamma^*(T, P_C) = \pi\sqrt{d}$. The last equality above is obtained by noting that $\|u\|_\infty$ is minimized over $u \in S^{d-1}$ when $|u_1| = \cdots = |u_d| = \sqrt{d}$.

Unlike what happens at spherically symmetric distributions, the gross error sensitivity, the contamination sensitivity, and the maximum bias function of Tukey's median at P_C are not independent of the dimension d, but rather are each proportional to $\sqrt{d}$. We conjecture, though, that this $\sqrt{d}$ rate may be the best possible rate at P_C for any translation equivariate location functional. The basis for this conjecture is that the scale $\sigma_u = \|u\|_1$ itself of the Cauchy distribution associated with the univariate projection $u'X$, with $u \in S^{d-1}$, can be of the order $\sqrt{d}$. That is, $\sup_{u \in S^{d-1}} \sigma_u = 1/\sqrt{d}$. A formal proof of the conjecture, however, appears to be rather challenging.

To conclude this subsection, we observe that the maximum contamination bias of the joint marginal medians T_M at P_{C} is given by

$$B(\varepsilon, T_M, P_C) = \sqrt{d}\tan\left(\frac{\pi}{2}\frac{\varepsilon}{1-\varepsilon}\right). \tag{40}$$

This is always less than the maximum bias of Tukey's median at P_C. The joint marginal medians, though, are not rotationally equivariant and one might expect them to perform best when the marginals are independent. That is, suppose we consider the distribution P_C^* associated with the random vector $X^* = Q'X$, where $Q = [q_1, \dots, q_d]$ is some $d \times d$ orthogonal matrix. The maximum contamination bias of Tukey's median is the same for all Q, whereas the maximum contamination bias of T_M depends on Q and can be shown to be

$$B(\varepsilon, T_M, P_C^*) = \left\{\sum_{i=1}^{d} \|q_i\|_1^2\right\}^{1/2} \tan\left(\frac{\pi}{2}\frac{\varepsilon}{1-\varepsilon}\right). \tag{41}$$

This quantity is smallest when Q is the identity matrix. It is largest when all the elements of Q are either $\pm 1/\sqrt{d}$, which is possible whenever $d = 2^k$ for some positive integer k. For such Q, $B(\varepsilon, T_M, P_C^*) = d\tan(\frac{\pi}{2}\frac{\varepsilon}{1-\varepsilon})$, which increases at a $\sqrt{d}$ rate faster than the maximum contamination bias of Tukey's median.

5.3. *Uniform distribution over a triangle.* To illustrate the results of Section 4 on a distribution which does not possess halfspace symmetry, we consider here the uniform distribution within the triangle $\triangle$ in $\Re^2$ having vertices $(-1, 0), (1, 0)$ and $(0, \sqrt{3})$, which we denote by $P_\triangle$. For this distribution, Rousseeuw and Ruts (1999) show that Tukey's depth at a point $x \in \Re^2$ in the region $G = \{x \mid -1 \le x_1 \le 1 \text{ and } 0 \le x_2 \le \frac{1-|x_1|}{\sqrt{3}}\}$ is given by $\pi(x; P) = \frac{2}{3}[\sqrt{3}(1-|x_1|)x_2 - x_2^2]$. The depth of other points inside the triangle $\triangle$ follows from the symmetry of the triangle. The contours of $\pi(x; P)$ and the region G are illustrated in Figure 3(a). The depth of points on or outside the triangle $\triangle$ is of course zero. The maximum depth $\Pi(P_\triangle) = \frac{4}{9}$ is uniquely obtained at $T(P_\triangle) = (0, \frac{1}{\sqrt{3}})$. [In this subsection, we depart from our convention of using translation to obtain $T(P) = 0$.]

From Theorem 4.1, we know that the maximum bias function $B(\varepsilon; T, P_\triangle)$ is bounded above by $\|L(\frac{\varepsilon}{1-\varepsilon}; P_\triangle)\|$ for $\varepsilon < \frac{4}{13}$. To evaluate this upper bound, we note that the convexity of the depth contours implies that the largest distance from the center of the triangle $(0, \frac{1}{\sqrt{3}})$ to a depth contour must occur either along the line segment from the center to $(0, 0)$ or along the line segment from the center to $(0, 1)$. After some lengthy but straightforward calculations, this gives

$$B(\varepsilon; T, P_\triangle) \le \begin{cases} \dfrac{\sqrt{3}}{6}\left(\sqrt{\dfrac{1+17\varepsilon}{1-\varepsilon}} - 1\right), & 0 \le \varepsilon \le \dfrac{20}{101}, \\[2ex] \dfrac{2\sqrt{3}}{3}\left(1 - \dfrac{1}{2}\sqrt{\dfrac{4-13\varepsilon}{1-\varepsilon}}\right), & \dfrac{20}{101} < \varepsilon < \dfrac{4}{13}. \end{cases} \tag{42}$$

The upper bound for the maximum bias function given in (42) is poor for larger values of ε since the breakdown point of Tukey's median at $P_\triangle$ is $\frac{1}{3}$. For larger

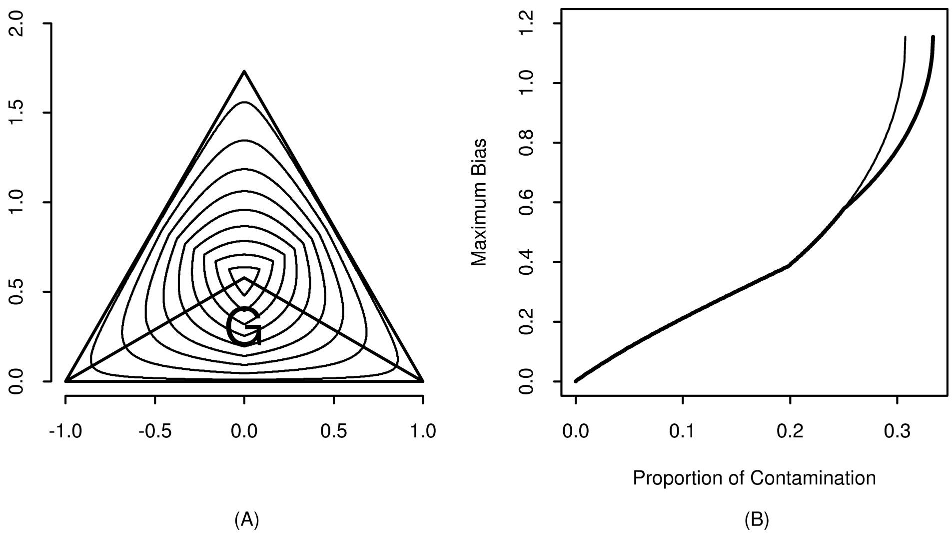

FIG. 3. (a) *The depth contours of* $P_\triangle$ *and the region* G. *The depth contours, from the outermost to the innermost contour, represent depths of* 0.01, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35 *and* 0.40, *respectively*. (b) *Upper bounds* (42) (*thin line*) *and* (43) (*heavy line*) *for* $B(\varepsilon; T, P_\triangle)$.

values of ε, a tighter upper bound can be obtained by applying Theorem 4.2, with $\delta(\varepsilon; P_\triangle)$ in Theorem 4.2 replaced $\delta_o(\varepsilon; P_\triangle) = \max\{0, \frac{4\varepsilon-1}{9(1-\varepsilon)}\}$. Again, after some lengthy but straighforward calculations, we obtain

$$B(\varepsilon; T, P_\triangle) \le \begin{cases} \dfrac{\sqrt{3}}{6}\left(\sqrt{\dfrac{1+17\varepsilon}{1-\varepsilon}} - 1\right), & 0 \le \varepsilon \le \dfrac{20}{101}, \\ \dfrac{2\sqrt{3}}{3}\left(1 - \dfrac{1}{2}\sqrt{\dfrac{4-13\varepsilon}{1-\varepsilon}}\right), & \dfrac{20}{101} < \varepsilon \le \dfrac{1}{4}, \\ \dfrac{2\sqrt{3}}{3}\left(1 - \dfrac{1}{2}\sqrt{\dfrac{3(1-3\varepsilon)}{1-\varepsilon}}\right), & \dfrac{1}{4} < \varepsilon < \dfrac{1}{3}. \end{cases} \tag{43}$$

Plots of the upper bounds (42) and (43) are given in Figure 3(b). An upper bound for the contamination sensitivity can be obtained by evaluating the derivative at zero for the upper bound in (42). This gives $\gamma^*(T, P_\triangle) \le \frac{3\sqrt{3}}{2}$.

APPENDIX: PROOFS AND TECHNICAL RESULTS

PROOF OF THEOREM 3.1. Since P is halfspace symmetric about the origin, $P\{H(u,0)\} = \frac{1}{2}$ for all $u \in S^{d-1}$. By considering half spaces which do not include x, we see that $\pi(0; P(\varepsilon, x)) = \frac{1-\varepsilon}{2}$ and that $\pi(y; P(\varepsilon, x)) < \frac{1-\varepsilon}{2}$ for any $y \notin \{\lambda x : 0 \le \lambda \le 1\}$. Therefore, $\mathrm{Med}(P(\varepsilon, x)) \subset \{\lambda x : 0 \le \lambda \le 1\}$.

Now for $y = \lambda x$ with $0 < \lambda < 1$, we can chose a sequence of vectors $u_i \in S^{d-1}$ such that the halfspaces $H(u_i, y)$ do not contain x but get arbitrarily close to x. Each element of the sequence $H(u_i, y)$ thus contains the origin with their boundaries being arbitrarily close to the origin. This implies $P(\varepsilon, x)\{H(u_i, y)\}$ gets arbitrarily close to $\frac{1-\varepsilon}{2}$, and so $\pi(y; P(\varepsilon, x)) \le \frac{1-\varepsilon}{2}$.

Consider first the case $x \notin L(\frac{\varepsilon}{1-\varepsilon}; P)$. For this case, $\pi(x; P(\varepsilon, x)) < \frac{1-\varepsilon}{2}$ and thus $\Pi(P(\varepsilon, x)) = \frac{1-\varepsilon}{2}$. Likewise for $y = \lambda x$ with $0 < \lambda < 1$, $\pi(y; P(\varepsilon, x)) < \frac{1-\varepsilon}{2}$ if and only if $y \notin L(\frac{\varepsilon}{1-\varepsilon}; P)$. Thus, $\mathrm{Med}(P(\varepsilon, x)) = L(\frac{\varepsilon}{1-\varepsilon}; P) \cap \{tx : t \ge 0\}$. The result for this case then follows from (9) or by noting that any affine equivariant preserving average over a line must give the midpoint of the line.

Consider next the case when x is on the boundary of $L(\frac{\varepsilon}{1-\varepsilon}; P)$. For this case, $\pi(x; P(\varepsilon, x)) = \frac{1-\varepsilon}{2}$ and so since the depth regions are convex we again have $\mathrm{Med}(P(\varepsilon, x)) = L(\frac{\varepsilon}{1-\varepsilon}; P) \cap \{tx : t \ge 0\}$. The result for this case then follows as before.

Finally, consider the case when x is in the interior of $L(\frac{\varepsilon}{1-\varepsilon}; P)$. For this case, $\pi(x; P(\varepsilon, x)) > (1-\varepsilon)(\frac{1}{2} - \frac{\varepsilon}{1-\varepsilon}) + \varepsilon = \frac{1-\varepsilon}{2}$. This implies that $\mathrm{Med}(P(\varepsilon, Q)) = x$ and so $T(P(\varepsilon, x)) = x$. $\square$

PROOF OF EQUATION (13). Consider first the spherical case, that is when $\Sigma = I_d$. For this case, the α-level sets $L(\alpha; P)$ are easily shown to be the interiors and boundaries of spheres centered at the origin and with radii $r_o(\alpha) = F_o^{-1}(\frac{1}{2} + \alpha)$. For the general elliptical case, the sets $L(\alpha; P)$ are obtained by considering affine transformations of the spherically symmetric case. $\square$

PROOF OF EQUATION (15). When $\Sigma = I_d$, P has density $g(x'x)$ and the marginal density of x_1 is given by $f_o(x_1)$. It then follows that $f_o(0) = \int_{\Re^{d-1}} g(x_2'x_2)\, dx_2$, where $x_2 \in \Re^{d-1}$. Application of the standard polar transformation to x_2 [see, e.g., Proposition 2.23 in Bilodeau and Brenner (1999)] yields (15) with ω_{d-1} being the surface area of the unit sphere in $\Re^{d-1}$. $\square$

PROOF OF BOUND (18). Let $y = r(\alpha, u)u$. Then by definiton $y \in L(\alpha; P)$ and so $\pi(y; P) \ge \frac{1}{2} - \alpha$. This implies $P\{H(-u, y)\} \ge \frac{1}{2} - \alpha$. However, $P\{H(-u, y)\} = 1 - F_u(u'y) = 1 - F_u(r(\alpha, u))$. Thus, $F_u(r(\alpha, u)) \le \frac{1}{2} + \alpha$ and (18) then follows. $\square$

PROOF OF THEOREM 3.4. We first show that for any Q, $\mathrm{Med}(P(\varepsilon, Q)) \subset L(\frac{\varepsilon}{1-\varepsilon}; P)$. To do so, we note that $\pi(0; P(\varepsilon, Q)) \ge \frac{1-\varepsilon}{2}$ and so $\Pi(P(\varepsilon, Q)) \ge \frac{1-\varepsilon}{2}$. Now $\pi(y; P(\varepsilon, Q)) \le (1-\varepsilon)\pi(y; P) + \varepsilon$, and so if $y \notin L(\frac{\varepsilon}{1-\varepsilon}; P)$, then $\pi(y; P) < \frac{1}{2} - \frac{\varepsilon}{1-\varepsilon}$ and, consequently, $\pi(y, P(\varepsilon, Q)) < \frac{1-\varepsilon}{2}$. Hence, if

$y \notin L(\frac{\varepsilon}{1-\varepsilon}; P)$, then $y \notin \mathrm{Med}(P(\varepsilon, Q))$, which implies $\mathrm{Med}(P(\varepsilon, Q)) \subset L(\frac{\varepsilon}{1-\varepsilon}; P)$. The results concerning the maximal bias function then follow since Theorem 3.1 states that if x is in the interior of $L(\frac{\varepsilon}{1-\varepsilon}; P)$ then $T(P(\varepsilon, x)) = x$.

The results on the maximum bias function imply that $\varepsilon^*(T; P) \geq \frac{1}{3}$. To show $\varepsilon^*(T; P) \leq \frac{1}{3}$ and hence equality, consider $P(\varepsilon, x)$ with $\varepsilon > \frac{1}{3}$. Now, by considering any half-plane that does not contain x, we see that $\pi(y; P(\varepsilon, x)) \leq \frac{1-\varepsilon}{2} < \varepsilon$ for any $y \neq x$. However, $\pi(x; P(\varepsilon, x)) \geq \varepsilon$ and so $T(P) = x$, which can be taken to be arbitrarily large. $\square$

PROOF OF THEOREMS 4.1 AND 4.2. To begin, note that $\pi(0; P(\varepsilon, Q)) \geq (1-\varepsilon)\Pi(P)$, which implies $\Pi(\varepsilon; P) \geq (1-\varepsilon)\Pi(P)$ and hence $\delta(\varepsilon; P) \geq 0$ and $\varepsilon_L(T; P) \geq \frac{\Pi(P)}{1+\Pi(P)}$. Thus, we see that Theorem 4.1 is a corollary to Theorem 4.2.

The proof of the bound on $B(\varepsilon; T, P)$ in Theorem 4.2 is similar to the first part of the proof of Theorem 3.4. First note that if $\varepsilon < \varepsilon_L(T; P)$ and we let $\alpha = \frac{\varepsilon}{1-\varepsilon} - \delta(\varepsilon; P)$, then $0 \leq \alpha < \Pi(P)$. Now $\pi(y; P(\varepsilon, Q)) \leq (1-\varepsilon)\pi(y; P) + \varepsilon$ and so if $y \notin L(\alpha; P)$, then $\pi(y; P) < \Pi(P) - \alpha$ and, consequently, $\pi(y; P(\varepsilon, Q)) < \Pi(\varepsilon; P)$. Hence, if $y \notin L(\alpha; P)$, then $y \notin \mathrm{Med}(P(\varepsilon, Q))$, which implies $\mathrm{Med}(P(\varepsilon, Q)) \subset L(\frac{\varepsilon}{1-\varepsilon}; P)$. This gives us the stated bound on $B(\varepsilon; T, P)$.

The bound on the maximum bias in Theorem 4.2 implies $\varepsilon^*(T; P) \geq \varepsilon_L(T; P)$. The inequality $\varepsilon_L(T; P) \geq \frac{1}{d+1}$ follows since from Lemma 6.3 in Donoho and Gasko (1992) we know that $\Pi(\varepsilon; P) \geq \frac{1}{d+1}$, and hence $\varepsilon_L(T; P) = \Pi(\varepsilon_L(T; P), P) \geq \frac{1}{d+1}$. The proof that $\varepsilon^*(T; P) \leq \frac{1}{3}$ is identical to its proof in Theorem 3.4. $\square$

Before giving the proof for Theorem 4.3, we need some results which characterize $\mathrm{Med}(P(\varepsilon, Q))$. As a consequence of these results, we also obtain some new results on the uniqueness of the deepest point. The uniqueness results are not needed for the proof of Theorem 4.3. We state them for completeness. Note that the second condition of the following theorem is satisfied whenever $\inf_{u \in S^{d-1}} f_u(0) > 0$.

THEOREM 5.1. *If P is absolutely continuous with respect to Lebesgue measure on $\Re^d$ and $P(0 \leq u'X \leq \delta) > 0$ for all $u \in S^{d-1}$ and $\delta > 0$, then*:

(a) $\mu \in \mathrm{Med}(P)$ *if and only if the set of all closed halfspaces* $H_\mu = \{H(u, \mu) : P\{H(u, \mu)\} = \pi(\mu; P)\}$ *covers* $\Re^d$.

(b) *If the set of interiors of H_μ, that is, the set of all open halfspaces* $\{\mathrm{int}(H(u, \mu)) : H(u, \mu) \in H_\mu\}$ *covers* $\Re^d/\mu$, *then* $\mathrm{Med}(P) = \{\mu\}$.

(c) *In* $\Re^2$, $\mathrm{Med}(P)$ *consists of a unique point*.

PROOF. Without loss of generality, we let $\mu = 0$.

(a) If H_0 covers $\Re^d$, then, for any $x \neq 0$, there is a half space $H(u,0) \in H_0$ such that $x \in H(u,0)$, and hence $\pi(x; P) \leq P\{H(u,x)\} \leq P\{H(u,0)\} = \pi(0; P)$. Thus, $0 \in \mathrm{Med}(P)$.

On the other hand, assume $0 \in \mathrm{Med}(P)$ and for any $v \in S^{d-1}$, let $v_n = v/n$. Then there is a sequence $u_n \in S^{d-1}$, such that $\pi(v_n; P) = P\{H(u_n, v_n)\} \leq \pi(0; P) \leq P\{H(u_n, 0)\}$. This implies that if ${u_n}'v_n > 0$, then $P(0 \leq {u_n}'X \leq {u_n}'v_n) = 0$, which contradicts the conditions of the theorem. We thus have ${u_n}'v_n \leq 0$ and hence ${u_n}'v \leq 0$. Now, since $u_n \in S^{d-1}$, there is a convergent subsequence u_m with say u_0 as its limit. For this u_0, we have ${u_0}'v \leq 0$ and consequently $v \in H(u_0, 0)$. Furthermore, we also have $P\{H(u_0,0)\} = \lim_{m\to\infty} P\{H(u_m, v_m)\} \leq \pi(0; P) \leq P\{H(u_0, 0)\}$, which implies $H(u_0,0) \in H_0$. Therefore S^{d-1} and hence $\Re^d$ is covered by H_0.

(b) By the conditions of part (b), for any $x \neq 0$, there is an $H(u,0) \in H_0$ such that $x \in \mathrm{int}\{H(u,0)\}$, or equivalently such that $u'x < 0$. By assumption, $P(u'x < u'X \leq 0) > 0$ and so we have $\pi(x; P) \leq P\{H(u,x)\} < P\{H(u,0)\} = \pi(0; P)$ and hence $\mathrm{Med}(P) = \{0\}$.

(c) Since without loss of generality we take $T(P) = 0$, we need to show that $\mathrm{Med}(P) = \{0\}$. If $\pi(0; P) = \frac{1}{2}$, then P must be halfspace symmetric about 0 (since any probability measure on a circle that gives mass $\frac{1}{2}$ to each closed half circle is symmetric and continuous). It readily follows that $H(u,0) \in H_0$ for any $u \in S^1$ and hence the set of interiors of H_0 covers $\Re^2$. The uniqueness of the median follows from part (b).

If $\pi(0; P) \neq \frac{1}{2}$, then $\pi(0; P) < \frac{1}{2}$ since P is absolutely continuous [see Donoho and Gasko (1992)]. By part (b), it suffices to prove that the set of interiors of H_0 covers $\Re^2/0$. Suppose this is not true, that is, suppose there exists an $x \neq 0$ which is not covered by the interiors of H_0. By part (a), however, we know H_0 covers $\Re^2$. Together, this implies that $H_0 = \{H(u,0), H(-u,0)\}$ with $u \perp x$, which in turn implies $\pi(0; P) = \frac{1}{2}$, a contradiction. □

PROOF OF THEOREM 4.3. By Theorem 5.1(a), since by assumption $T(P) = 0$ and hence $0 \in \mathrm{Med}(P)$, we know that H_0 covers $\Re^d$. We can choose a finite subcovering, say $\{H(u_i, 0) : i = 1, \dots, k\}$. For each u_i and $\varepsilon > 0$, define $a_{i,\varepsilon} > 0$ such that $P\{H(u_i, -a_{i,\varepsilon}u_i)\} = \pi(0; P) - \frac{\varepsilon}{1-\varepsilon}$.

We begin by proving the claim that $L(\frac{\varepsilon}{1-\varepsilon}; P) \subset \bigcap_{i=1,\dots,k} H(-u_i, -a_i u_i)$. To do so, note that if $y \notin \bigcap_{i=1,\dots,k} H(-u_i, -a_{i,\varepsilon}u_i)$, then $u_i' y < -a_{i,\varepsilon}$ for some $i = 1, \dots, k$. For such i, we then have $P\{H(u_i, y)\} < P\{H(u_i, -a_{i,\varepsilon}u_i)\} = \pi(0; P) - \varepsilon/(1-\varepsilon)$. This implies $y \notin L(\frac{\varepsilon}{1-\varepsilon}; P)$, and so the claim follows.

Combining this claim with Theorem 4.1, we obtain $B(\varepsilon : T, P) \leq \| \bigcap_{i=1,\dots,k} H(-u_i, -a_i u_i)\|$. Now, $\bigcap_{i=1,\dots,k} H(-u_i, -a_i u_i)$ is a polyhedron. Moreover, $a_{i,\varepsilon} = F_{u_i}^{-1}(\pi(0; P)) - F_{u_i}^{-1}(\pi(0; P) - \varepsilon/(1-\varepsilon))$ and so $\lim_{\varepsilon\to 0} \frac{a_{i,\varepsilon}}{\varepsilon} = \{f_{u_i}(0)\}^{-1}$. Hence, $\limsup_{\varepsilon\to 0} \| \bigcap_{i=1,\dots,k} H(-u_i, -a_{i,\varepsilon}u_i)\|/\varepsilon < \infty$, which completes the proof. □

REFERENCES

ADROVER, J. and YOHAI, V. (2002). Projection estimates of multivariate location. *Ann. Statist.* **30** 1757–1778.

BAI, Z.-D. and HE, X. (1999). Asymptotic distributions of the maximal depth estimators for regression and multivariate location. *Ann. Statist.* **27** 1616–1637.

BILODEAU, M. and BRENNER, D. (1999). *Theory of Multivariate Statistics*. Springer, New York.

CHEN, Z. (1995). Robustness of the half space median. *J. Statist. Plann. Inference* **46** 175–181.

CROUX, C., HAESBROECK, G. and ROUSSEEUW, P. J. (2001). Location adjustment for the minimum volume ellipsoid estimator. Technical report, Free Univ. Brussels (ULB). Available at http://homepages.ulb.ac.be/~ccroux/public.htm.

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

DONOHO, D. L. and GASKO, M. (1992). Breakdown properties of location estimates based on half-space depth and projected outlyingness. *Ann. Statist.* **20** 1803–1827.

HAMPEL, F. R., RONCHETTI, E. Z., ROUSSEEUW, P. J. and STAHEL, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York.

HE, X. and SIMPSON, D. G. (1993). Lower bounds for contamination bias: Globally minimax versus locally linear estimation. *Ann. Statist.* **21** 314–337.

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

KENT, J. T. and TYLER, D. E. (1996). Constrained M-estimation for multivariate location and scatter. *Ann. Statist.* **24** 1346–1370.

LOPUHAÄ, H. P. (1989). On the relation between S-estimators and M-estimators of multivariate location and covariance. *Ann. Statist.* **17** 1662–1683.

LOPUHAÄ, H. P. (1992). Highly efficient estimators of multivariate location with high breakdown point. *Ann. Statist.* **20** 398–413.

MARDIA, K. V. (1972). *Statistics of Directional Data*. Academic Press, New York.

MARONNA, R. A. (1976). Robust M-estimators of multivariate location and scatter. *Ann. Statist.* **4** 51–67.

MARTIN, R. D., YOHAI, V. J. and ZAMAR, R. H. (1989). Min-max bias robust regression. *Ann. Statist.* **17** 1608–1630.

NOLAN, D. (1999). On min-max majority and deepest points. *Statist. Probab. Lett.* **43** 325–334.

RANDLES, R. H. (2000). A simpler, affine-invariant multivariate distribution-free sign test. *J. Amer. Statist. Assoc.* **95** 1263–1268.

ROUSSEEUW, P. J. and RUTS, I. (1999). The depth function of a population distribution. *Metrika* **49** 213–244.

SMALL, C. G. (1987). Measures of centrality for multivariate and directional distributions. *Canad. J. Statist.* **15** 31–39.

SMALL, C. G. (1990). A survey of multidimensional medians. *Internat. Statist. Rev.* **58** 263–277.

TUKEY, J. W. (1975). Mathematics and the picturing of data. In *Proc. International Congress of Mathematicians, Vancouver 1974* (R. D. James, ed.) **2** 523–531. Canadian Math. Congress, Montreal.

WATSON, G. S. (1983). *Statistics on Spheres*. Wiley, New York.

ZUO, Y. and SERFLING, R. (2000). General notions of statistical depth function. *Ann. Statist.* **28** 461–482.

DEPARTMENT OF MATHEMATICS
WILLIAM PATERSON UNIVERSITY
WAYNE, NEW JERSEY 07470
E-MAIL: ChenZ@wpunj.edu

DEPARTMENT OF STATISTICS
RUTGERS UNIVERSITY
110 FRELINGHUYSEN ROAD
PISCATAWAY, NEW JERSEY 08854-8019
E-MAIL: dtyler@rci.rutgers.edu