

5-1-2004

# Multivariate Location: Robust Estimators And Inference

Rand R. Wilcox

University of Southern California, [rwilcox@usc.edu](mailto:rwilcox@usc.edu)

H. J. Keselman

University of Manitoba, [kesel@ms.umanitoba.ca](mailto:kesel@ms.umanitoba.ca)

Follow this and additional works at: <http://digitalcommons.wayne.edu/jmasm>

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

## Recommended Citation

Wilcox, Rand R. and Keselman, H. J. (2004) "Multivariate Location: Robust Estimators And Inference," *Journal of Modern Applied Statistical Methods*: Vol. 3 : Iss. 1 , Article 2.

DOI: [10.22237/jmasm/1083369720](https://doi.org/10.22237/jmasm/1083369720)

Available at: <http://digitalcommons.wayne.edu/jmasm/vol3/iss1/2>

This Invited Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized editor of DigitalCommons@WayneState.

## INVITED ARTICLES

### Multivariate Location: Robust Estimators And Inference

Rand R. Wilcox  
Department of  
Psychology  
University of Southern  
California



H. J. Keselman  
Department of  
Psychology  
University of Manitoba



---

The sample mean can have poor efficiency relative to various alternative estimators under arbitrarily small departures from normality. In the multivariate case, (affine equivariant) estimators have been proposed for dealing with this problem, but a comparison of various estimators by Massé and Plante (2003) indicated that the small-sample efficiency of some recently derived methods is rather poor. This article reports that a skipped mean, where outliers are removed via a projection-type outlier detection method, is found to be more satisfactory. The more obvious method for computing a confidence region based on the skipped estimator (using a slight modification of the method in Liu & Singh, 1997) is found to be unsatisfactory except in the bivariate case, at least when the sample size is small. A much more effective method is to use the Bonferroni inequality in conjunction with a standard percentile bootstrap technique applied to the marginal distributions.

Keywords: Outlier detection; Tukey's halfspace depth, skipped estimators, outlier-projection estimator

---

#### Introduction

A fundamental problem is estimating a measure of location associated with some multivariate distribution and then computing a confidence region based on the estimator used. Of course, the sample mean performs well under normality based on various well-known criteria. However, from an applied point of view, there are compelling reasons to consider alternative measures of location. One has to do with the effects of outliers on efficiency. Tukey (1960) predicted that outliers are common in applied

work, and modern outlier detection methods indicate that this is indeed the case (outlier detection rules based on the mean and usual covariance matrix, in conjunction with a Mahalanobis distance, are well known to be unsatisfactory; see for example, Rousseeuw & Leroy, 1987).

Moreover, arbitrarily small departures from normality (based on any of several metrics for comparing distributions) can result in outliers commonly appearing in a random sample which in turn can mean poor efficiency when using the sample mean. Another concern is that when sampling from an asymmetric distribution, the population mean can poorly reflect what is typical.

In the univariate case, many alternatives to the sample mean have been proposed (e.g., Andrews et al., 1972). Several maintain relatively high accuracy under normality relative to the sample mean and have high efficiency in situations where the sample mean performs poorly. Simultaneously, inferential (hypothesis testing) methods have been found that perform

---

Rand R. Wilcox (rwilcox@usc.edu) is a Professor of Psychology at the University of Southern California and H. J. Keselman (kesel@ms.umanitoba.ca) is Professor of Psychology at the University of Manitoba. Work on this project was supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

well over a broad range of situations, including general conditions where methods based on means perform in an unsatisfactory manner (e.g., Wilcox, 1997; 2003).

As for the multivariate case, one could of course simply apply univariate estimators to the marginal distributions, but it is known that usually this does not satisfy a criterion that is frequently imposed. To elaborate, first consider the univariate case, let  $X$  be any random variable having distribution  $F$ , let  $T(X_1, \dots, X_n)$  be some statistic based on the random sample  $X_1, \dots, X_n$ , and let  $a$  and  $b$  be any two constants. Then for  $T(X_1, \dots, X_n)$  to qualify as a measure of location, a minimum requirement typically imposed is that

$$T[a(X_1, \dots, X_n) + b] = aT(X_1, \dots, X_n) + b$$

(e.g., Staudte & Sheather, 1990). The usual population mean and median satisfy this requirement as do many other robust measures of location. This requirement says, for example, that given a typical measure of temperature in Fahrenheit, if converted to Celsius, the typical measure should be transformed in the obvious way.

Now consider the case where  $X$  is any  $p$ -variate random variable,  $A$  is any nonsingular square matrix, and  $B$  is a vector having length  $p$ . Then  $T(X_1, \dots, X_n)$  is said to be an affine equivariant measure of location if

$$T(X_1A, \dots, X_nA + B) = T(X_1, \dots, X_n)A + B. \quad (1)$$

So the measure of location is transformed properly under rotations as well as changes in scale and shifts in the possible values of  $X$ . There are many robust affine equivariant measures of location in the univariate case, but typically, if they are applied to the marginal distributions in the multivariate case, they are no longer affine equivariant. For example, the marginal medians are not affine equivariant as noted by Donoho and Gasko (1992). So a general goal has been to search for affine equivariant location estimators in the multivariate case that guard against the deleterious effects of outliers.

One of the earliest affine equivariant estimators that guards against outliers was proposed by Rousseeuw and Leroy (1987) and is known as the minimum volume ellipsoid (MVE) estimator. It begins by searching for the ellipsoid containing half of the data that has the smallest volume. If the sample mean is computed based on this half of the data, ignoring the other half, it is evident that it guards against outliers, but efficiency is poor compared to the usual sample mean when sampling from a normal distribution. More recently, Rousseeuw and van Driessen (1999) argued that the MVE estimator be replaced by the minimum covariance determinant (MCD) estimator which searches for a subset of half of the data having the smallest generalized variance. But like the MVE estimator, efficiency is low when sampling from a multivariate normal distribution.

Donoho and Gasko (1992) studied a multivariate location estimator that is based in part on Tukey's notion of halfspace depth. Their approach is of direct interest in this article and details are given later in the paper. But before continuing, a rough outline of their strategy helps. The basic idea is to quantify how deeply each point  $X_i$  is nested within the cloud of data, and then eliminate a fixed proportion of those points that are not deeply nested.

That is, use the centrally located data to estimate a measure of location and ignore the data on the edges of the data cloud. In the univariate case, their estimator reduces to a trimmed mean which is known to have many practical advantages. In particular, a 20% trimmed mean (where the largest 20% and the smallest 20% of the observed values are trimmed, and the average of the remaining data is used) maintains reasonably high efficiency under normality (e.g., Rosenberger & Gasko, 1987). This raises the issue of whether a similar amount of trimming performs well when working with multivariate data, and it is found that this is not the case.

Yet another approach was recently proposed by Liu, Parelius and Singh (1999) and represents a generalization of the method studied by Donoho and Gasko (1992). One difference is that Liu et al. consider a wider choice of methods for measuring the depth of a point

within a data cloud. The particular generalization of a univariate trimmed mean described by Liu et al. (1999, pp. 795-796) was considered in this study, but it did not correct the problems with the Donoho and Gasko trimmed mean described later in this article. So for brevity, the complex computational details of their method are not described here.

One more general approach is to first search for outliers using some affine equivariant method, roughly meaning that if the data are transformed as indicated by the left side of equation 1, outliers before transforming the data remain outliers after transformation. Next, eliminate any outliers that are found and simply compute the mean of the remaining data. In the univariate case, this general strategy yields what is known as a skipped estimator, so the term is used here.

The focus here is on one particular outlier detection method for reasons discussed later in the paper. The choice of method is not arbitrary, but it is stressed that alternative outlier detection techniques might be found to have practical value in future studies. It is noted that Massé and Plante (2003) compared the efficiency of several multivariate estimators and found all of the affine equivariant estimators to have relatively poor efficiency under normality. The skipped estimator studied here corrects that problem.

#### The Estimators Studied

This section provides a formal description of the six estimators considered. Four of the estimators belong to the class of generalized trimmed means studied by Donoho and Gasko (1992); four different amounts of trimming are considered. Results in Massé and Plante (2003) indicate that these estimators can be unsatisfactory, and previous results, based on other distributions and criteria, support their conclusions. The fifth estimator is based on removing outliers with a projection-type method and averaging the values that remain, and the sixth is the usual median of the marginal distributions. Although this last estimator is not affine equivariant, it is included with the goal of adding perspective on the expected accuracy of the other estimators considered.

#### The Donoho-Gasko Trimmed Mean

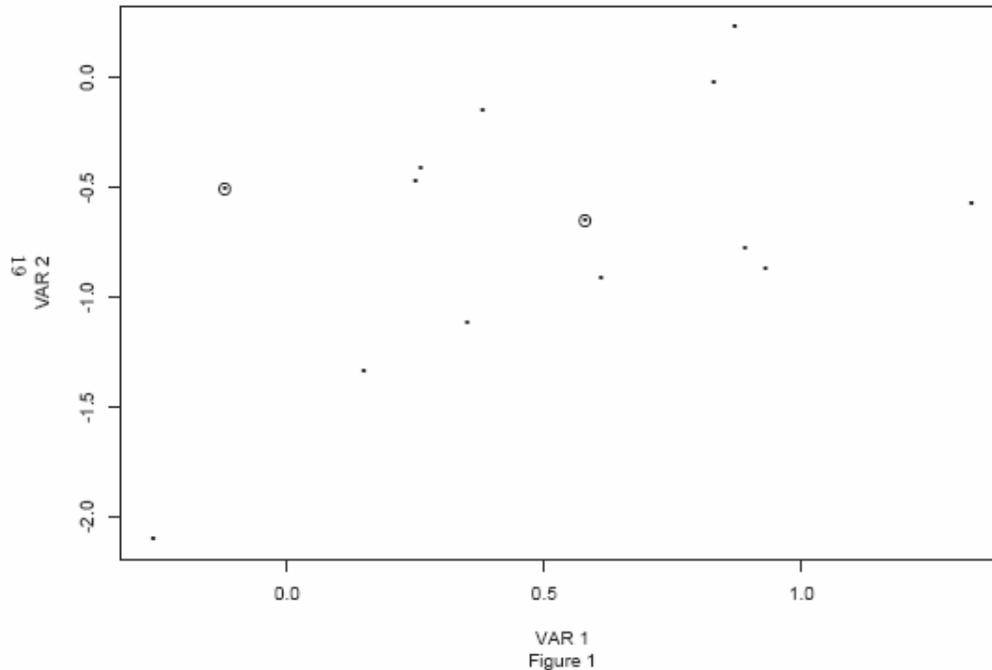
The Donoho and Gasko (1992) estimator is based on Tukey's notion of halfspace depth, which represents an approach to generalizing the notion of ranks to multivariate data. An important feature of Tukey's depth is that no assumptions are made about the distribution from which observations are randomly sampled. In particular, it is not assumed that the distribution is elliptical.

A formal definition of Tukey's depth is relegated to an appendix. To provide some intuitive sense of Tukey's strategy we duplicate a description found in Wilcox (in press). Look at Figure 1 which shows a scatterplot of electroencephalographic (EEG) measures taken at two sites in the brain. These data are from Raine, Buchsbaum and LaCasse (1997) where the general goal was to investigate brain abnormalities among murderers.

Consider the left most point indicated by a circle and imagine any line going through this point. Any line forms what are called two halfspaces. The points on or above a line form a closed halfspace, and the same is true for all of the points on or below the line. Because the left point indicated by the circle is located on the edge of the scatterplot, it is evident that a line can be drawn through this point so that it is the only point in one of the closed halfspaces. Now consider the right circle. Because it is more deeply nested within the scatterplot, a relatively large proportion of the scatterplot will be on or above any line drawn through this point, and a relatively large proportion will be on or below the line as well.

For any line  $L$  drawn through a point, consider the proportion of points on or above this line, as well as on or below this line, and let  $P_L$  be the smaller of these two proportions. Then Tukey's depth is the smallest  $P_L$  value among all lines  $L$ . For  $p$ -variate data (where  $L$  becomes a plane), the maximum depth among a scatterplot of points can be as high as  $1/2$  or as low as  $1/(p+1)$  (Donoho & Gasko, 1992). So for bivariate data ( $p=2$ ), if the depth for every point were computed, it is possible that the

Figure 1. EEG measures used to illustrate Tukey's notion of depth.



largest depth would not exceed  $1/3$ , but it could be as high as  $.5$ .

Tukey's notion of depth can be computed exactly in the bivariate case (Rousseeuw & Ruts, 1996). In Figure 1, there are  $n = 14$  points, and the halfspace depth for the left circle is  $1/14$ . For the right circle, the halfspace depth is  $5/14$ . For  $p$ -variate data,  $p > 2$ , Rousseeuw and Struyf (1998) describe an approximation of Tukey's depth which is used here.

The Donoho-Gasko analog of the  $\gamma$ -trimmed mean,  $\hat{\xi}_\gamma$ , is the average of all points which are at least  $\gamma$  deep in the sample. That is, points having depth less than  $\gamma$  are trimmed and the mean of the remaining points is computed. For example, suppose  $\gamma = .1$  and consider again the data in Figure 1. There are four points that have a depth less than  $.1$  so the  $.1$  trimmed mean

is the mean after these four points are eliminated. Because the maximum depth is not necessarily  $.5$ , a generalization of the median, often called Tukey's median, is taken to be the mean of the points having the maximum depth. In Figure 1, the maximum depth is  $.357$  which corresponds to only one point:  $(.58, -.65)$ .

#### A Skipped Estimator

As previously indicated, a skipped estimator is the sample mean of the data after outliers have been removed. A practical problem is not finding a reasonable outlier detection method for multivariate data, but rather choosing a method from among the many that have been proposed. Rousseeuw, Ruts and Tukey (1999) suggest a method based on the notion of halfspace depth. They focus mainly on the bivariate case, but in principle the method can be used when  $p > 2$ ; also see Liu et al. (1999) as well as Romanazzi (1997).

An approach based on convex hull peeling is discussed by Zani, Riani and Corbellini (1998) but is known to be somewhat less robust than halfspace depth as shown by Donoho and Gasko (1992). Another approach, that has been studied extensively, is related to the strategy behind the MVE and MCD estimators previously described. That is, find the ellipsoid with the smallest volume or smallest covariance determinant that encompasses at least half of the data, and use the corresponding mean and covariance matrix to detect outliers. (See, for example, Davies, 1987; Fung, 1993; Hampel, Ronchetti, Rousseeuw & Stahel, 1986; Rousseeuw & Leroy, 1987; Rousseeuw & van Zomeren, 1990; Rousseeuw & van Driessen, 1999; Tyler, 1991; For additional references, see Peña & Prieto, 2001; cf. Woodruff & Rocke, 1994.)

The main article for detecting outliers based on the minimum volume ellipsoid (MVE) estimator is Rousseeuw and van Zomeren (1990). Rocke and Woodruff (1996) describe a method that uses the MVE and MCD estimators as starting values for computing estimators of location and scatter. Poon, Lew and Poon (2000) suggest a method based in part on a Mahalanobis distance, and yet another approach was recently proposed by Viljoen and Venter (2002).

One more strategy, as suggested by Stahel (1981) and Donoho (1982), is motivated by the fact that each outlier among a multivariate sample must be an extreme point based on some projection of the data. Adopting this view, Peña and Prieto (2001) focus on how far points are from the usual sample mean, and they suggest how to choose interesting projections based on the estimated kurtosis coefficient of the projected observations.

This study uses a projection-type method for detecting outliers for reasons to be described and because software is easily written to perform the calculations. To reduce the number of projections considered, the strategy used by Peña and Prieto (2001) is used where attention is focused on how far a point is from the center of the data. The idea is that by projecting points onto a line that passes through the center of the data, the distances between points on the projected line can be combined

with known properties of univariate outlier detection methods in a manner that are advantageous for the problem at hand.

But rather than use the sample mean, as was done by Peña and Prieto, the Donoho and Gasko (1992) multivariate median estimator  $\hat{\xi}_m$ , is used instead. Another difference is that  $n$  projections are considered. In contrast, with  $p$ -variate data, Peña and Prieto search for  $2p$  projections instead.

To briefly elaborate, it is noted that the so-called outside rate per observation for an outlier detection method refers to the proportion of points declared outliers based on a sample of size  $n$ . When searching for an estimator that performs nearly as well as the sample mean under normality, it seems clear that the outside rate per observation should be reasonably low when sampling is from a multivariate normal distribution. Known results on univariate outlier detection methods, suggest how to control the outside rate per observation when considering projections, so this strategy is used in the current study.

The details of the method used here are as follows. Fix  $i$  and for the point  $X_i$ , project all  $n$  points onto the line connecting  $\hat{\xi}_m$  and  $X_i$  and let  $D_j$  be the distance between  $\hat{\xi}_m$  and  $X_j$  based on this projection. More formally, let

$$A_i = X_i - \hat{\xi}_m,$$

$$B_j = X_j - \hat{\xi}_m,$$

where both  $A_i$  and  $B_j$  are column vectors having length  $p$ , and let

$$C_j = \frac{A_i' B_j}{B_j' B_j} B_j,$$

$j=1, \dots, n$ . Then when projecting the points onto the line between  $X_i$  and  $\hat{\xi}_m$ , the distance of the  $j$ th point from  $\hat{\xi}_m$  is

$$D_j = \|C_j\|,$$

where  $\|C_j\|$  is the Euclidean norm of the vector  $C_j$ .

Next, a boxplot rule for detecting outliers is applied to the  $D_j$  values, but rather than the standard rule, a modification that has close similarities to one used by Carling (2000) is employed. Let  $l = [n/4 + 5/12]$ , where  $[.]$  is the greatest integer function, and let

$$h = \frac{n}{4} + \frac{5}{12} - l.$$

Let  $D_{(1)} \leq \dots \leq D_{(n)}$  be the  $n$  distances written in ascending order. The so-called ideal fourths associated with the  $D_j$  values are

$$q_1 = (1-h)D_{(j)} + hD_{(j+1)}$$

and

$$q_2 = (1-h)X_{(k)} + hX_{(k-1)}.$$

Then the  $j$ th point is declared an outlier if

$$D_j > M_D + \sqrt{\chi_{.95,p}^2} (q_2 - q_1), \quad (2)$$

where  $M_D$  is the usual sample median based on the  $D_j$  values and  $\chi_{.95,p}^2$  is the .95 quantile of a chi-squared distribution with  $p$  degrees of freedom (cf. Rousseeuw & van Zomeren, 1999).

The process just described is for a single projection; for fixed  $i$ , points are projected onto the line connecting  $X_i$  to  $\hat{\xi}_m$ . Repeating this process for each  $i$ ,  $i=1, \dots, n$ , a point is declared an outlier if for any of these projections, it satisfies equation (2). Removing any outliers found by equation (2), and averaging the values that remain, will be called the OP (outlier-projection) estimator and denoted by  $\hat{\xi}_{op}$ .

A simple and seemingly desirable modification of the method just described is to replace the interquartile range ( $q_2 - q_1$ ) with the median absolute deviation (MAD) measure of

scale based on the  $D_j$  values. So here, MAD is the median of the values

$$|D_1 - M_D|, \dots, |D_n - M_D|.$$

Then the  $j$ th point is declared an outlier if

$$D_j > M_D + \sqrt{\chi_{.95,p}^2} \left( \frac{\text{MAD}}{.6745} \right), \quad (3)$$

where the constant .6745 is typically used because under normality,  $\text{MAD}/.6745$  estimates the standard deviation. (Equation 3 represents an approximation of the method given by equation 1.3 in Donoho & Gasko, 1992.) One appealing feature of MAD is that it has a higher finite sample breakdown point versus the interquartile range, where the finite sample breakdown point of an estimator refers to the minimum proportion of points that must be altered to make the value of a statistic arbitrarily small or large. MAD has a finite sample breakdown point of approximately .5, while for the interquartile range it is only .25.

In this study, however, the focus is on using a projection-type method in conjunction with the interquartile range, rather than MAD. The reason stems from the outside rate per observation,  $p_n$ . As previously suggested, to maintain relatively high accuracy under normality when using a skipped estimator, the outside rate per observation should be reasonably close to zero.

It is common to search for a method with a rate approximately equal to .05; this usually provides good efficiency under normality. A negative feature of equation (3) is that  $p_n$  appears to be considerably less stable as a function of  $n$ . In the bivariate case, for example, it is approximately .09 with  $n=10$  and drops below .02 as  $n$  increases. So the relative accuracy of the corresponding skipped estimator varies with  $n$ . For the same situations,  $p_n$  based on equation 2 ranged between .043 and .038. So the approach based on equation 3 is not pursued here.

To further elaborate on why the MVE outlier detection method was discarded, it is

noted that under normality, and when all variables are independent, its outside rate per observation is approximately .05, but when there is dependence, the rate can be considerably higher (Wilcox, 2003). The result is that if points declared outliers are removed, and the sample mean of the remaining points is computed, efficiency remains relatively high under independence, but it can be relatively low otherwise, so this approach was abandoned. If the MVE measures of location and scatter are replaced by the MCD estimators previously mentioned, again general situations were found where efficiency is poor under normality, which was not surprising because in these situations the outside rate per observation was even higher than was found for the MVE method.

The outside rate per observation for many outlier detection methods has not been studied and addressing this issue goes beyond the scope of this article. So, of course, some variation of the skipped estimator studied here might give improved results in some sense, but this remains to be determined.

Confidence Region

Given that location is estimated using  $\hat{\xi}_{op}$ , how should a confidence region for  $\xi$ , the parameter estimated by  $\hat{\xi}_{op}$ , be computed? The initial strategy was to use the bootstrap method in Liu and Singh (1997). A direct application of their method, or some slight variation of it, has been found to perform well for a wide range of problems (Wilcox, 2003). Here, however, this approach was found to be unsatisfactory and was eventually abandoned. A much more satisfactory approach, in simulations, is to proceed as follows.

Let  $\xi_j$  represent the  $j$ th element of the vector  $\xi$ ,  $j = 1, \dots, p$ . Let  $X_{ij}$ ,  $i = 1, \dots, n; j = 1, \dots, p$ , represent a random sample from some  $p$ -variate distribution. Generate a bootstrap sample by resampling with replacement  $n$  rows from the  $n$  by  $p$  matrix corresponding to  $X_{ij}$ , and denote this bootstrap sample by  $X_{ij}^*$ . Let  $\hat{\xi}^* = (\hat{\xi}_1^*, \dots, \hat{\xi}_p^*)$  represent the OP estimate of  $\xi$  based on this bootstrap sample.

Repeat this process  $B$  times and let  $\hat{\xi}_{jb}^*$  be the estimate of  $\xi_j$  based on the  $b$ th bootstrap sample,  $b = 1, \dots, B$ . Then from basic principles (e.g., Efron & Tibshirani, 1993), an approximate  $1 - \alpha$  confidence interval for  $\xi_j$  is given by  $(\hat{\xi}_{j(l+1)}^*, \hat{\xi}_{j(u)}^*)$ , where for fixed  $j$   $\hat{\xi}_{j(1)}^* \leq \dots \leq \hat{\xi}_{j(B)}^*$  are the  $\hat{\xi}_{jb}^*$  values written in ascending order,  $l = \alpha B / 2$ , rounded to the nearest integer, and  $u = B - l$ . So, to obtain an approximate  $1 - \alpha$  confidence region for  $\xi$ , a simple strategy is to apply the Bonferroni inequality and compute an  $\alpha/p$  confidence interval for  $\xi_j$  using the method just described.

Said another way, to test  $H_0 : \xi_j = \xi_{0j}$ ,  $\xi_{0j}$  given, let  $p_j^*$  be the probability that  $\hat{\xi}_j^*$  is less than  $\xi_{0j}$ . From Liu and Singh (1997), for fixed  $j$ ,  $p_j^*$  has, asymptotically, a uniform distribution. Although  $p_j^*$  is not known, it is readily estimated from the data with

$$\hat{p}^* = \frac{A}{B}, \tag{4}$$

where  $A$  is the number of bootstrap samples with  $\hat{\xi}_{jb}^* < \xi_{0j}$ . Then  $2\hat{p}_m^*$  is the estimated p-value, where

$$\hat{p}_m^* = \min(\hat{p}^*, 1 - \hat{p}^*). \tag{5}$$

So for fixed  $j$ , reject at the  $\alpha$  level if  $2\hat{p}_m^* \leq \alpha$ . To control the familywise error rate (the probability of at least one Type I error) via the Bonferroni inequality when testing all  $p$  hypotheses, reject if  $2\hat{p}_m^* \leq \alpha / p$ .

Methodology

Simulations were used to check both the accuracy of the estimators considered, plus the actual probability coverage when using the method just discussed. Accuracy was measured using the sum of the squared standard errors

associated with the  $p$  estimators used to estimate the  $p$  parameters. For the OP estimator, again let  $\hat{\xi}_j$  be the OP estimate of  $\xi_j$ , and let  $\bar{X}_j$  be the usual sample mean corresponding to the  $j$ th marginal distribution. Then the accuracy of the OP estimator, relative to the sample mean, is measured by  $R^2$ , the sum of the squared standard errors of the estimators associated with  $\hat{\xi}_j, j=1, \dots, p$ , divided by the sum of the squared standard errors associated with the sample means,  $\bar{X}_j$ . When dealing with other estimators, the  $\hat{\xi}_j$  were replaced with the relevant estimator.

Observations were generated where the marginal distributions have a *g-and-h* distribution (Hoaglin, 1985) which includes normal distributions as a special case. When dealing with accuracy, the focus was on  $p=4$  (but when computing a confidence region, both  $p=2$  and 4 were considered). More precisely, observations  $Z_{ij}=1, \dots, n; j=1, \dots, p$

were initially generated from a multivariate normal distribution having correlation  $\rho$ , then the marginal distributions were transformed to

$$X_{ij} = \frac{\exp(g Z_{ij}) - 1}{g} \exp\left(\frac{h Z_{ij}^2}{2}\right)$$

when both  $g$  and  $h$  were non-zero. When  $g$  was zero

$$X_{ij} = Z_{ij} \exp\left(\frac{h Z_{ij}^2}{2}\right),$$

where  $g$  and  $h$  are parameters that determine the third and fourth moments. Here,  $\rho=0$  and  $.7$  are considered. The four (marginal) *g-and-h* distributions used here were the standard normal ( $g=h=0$ ), a symmetric heavy-tailed distribution ( $g=0, h=.5$ ), an asymmetric distribution with relatively light tails ( $g=.5, h=0$ ), and an asymmetric distribution with heavy tails ( $g=h=.5$ ). Also, when dealing with accuracy, simulations were run with  $h=1$ . This latter case might be viewed as

an extreme departure from normality, but it was considered anyway to see whether any of the estimators performs poorly when sampling from a sufficiently heavy-tailed distribution.

Table 1 shows the theoretical skewness ( $\kappa_1$ ) and kurtosis ( $\kappa_2$ ) values for each distribution considered. When  $g > 0$  and  $h > 1/k$ ,  $E(X^k)$  is not defined and the corresponding entry in Table 1 is left blank. Additional properties of the *g-and-h* distribution are summarized by Hoaglin (1985).

Table 1. Some properties of the *g-and-h* distribution.

$g$	$h$	$\kappa_1$	$\kappa_2$	$\hat{\kappa}_1$	$\hat{\kappa}_2$
0.0	0.0	0.00	3.0	0.00	3.0
0.0	0.5	0.00	—	0.00	11,896.2
0.5	0.0	1.75	8.9	1.81	9.7
0.5	0.5	—	—	120.10	18,393.6

A possible objection to Table 1 when performing simulations is that the distribution of observations generated on a computer does not always have the theoretical skewness and kurtosis values shown. The reason is that computer observations come from a bounded interval, so the skewness and kurtosis of the distribution will be finite, even when in theory it should be infinite. Accordingly, Table 1 also reports the estimated skewness ( $\hat{\kappa}_1$ ) and kurtosis ( $\hat{\kappa}_2$ ) values based on simulations with 10,000 replications.

Table 2 shows estimates of  $R$  based on 5,000 replications, where the first three estimators are the Donoho-Gasko trimmed mean with 10%, 15% and 20% trimming, DGM is the Donoho-Gasko median, OP is the outlier-projection estimator, and  $M$  is the usual median. Note that with 20% trimming, accuracy is relatively poor when sampling from a normal distribution ( $g=h=0$ ).

Table 2. Values of  $R$  (Accuracy),  $n = 40$

$g$	$h$	$\rho$	$\gamma = .10$	$\gamma = .15$	$\gamma = .20$	DGM	OP	M
0.0	0.0	0.0	0.73	0.62	0.50	0.45	0.92	0.81
0.0	0.5	0.0	5.99	5.92	5.40	4.11	6.25	8.48
0.0	1.0	0.0	4660.21	5764.79	5911.29	4643.16	5452.35	10820.14
0.0	0.0	0.7	0.80	0.71	0.61	0.48	0.95	0.44
0.0	0.5	0.7	4.74	4.76	4.50	3.20	4.64	5.44
0.0	1.0	0.7	1082.56	1300.44	1336.63	1005.24	1091.68	1760.98
0.5	0.0	0.0	0.79	0.69	0.54	0.49	0.99	0.99
0.5	0.5	0.0	13.01	12.78	11.82	8.91	14.95	20.66
0.5	1.0	0.0	1908.75	2413.39	2472.07	1852.97	2519.04	4887.50
0.5	0.0	0.7	0.94	0.86	0.69	0.53	1.05	0.99
0.5	0.5	0.7	17.79	18.05	17.22	11.34	17.42	20.66
0.5	1.0	0.7	3005.56	3652.36	3660.06	29996.40	4887.42	4887.40

This is in sharp contrast with the univariate case where a 20% trimmed mean performs reasonably well (e.g., Rosenberger & Gasko, 1983; Wilcox, 1997). Under normality, with  $\rho = .7$ , the median performs rather poorly, but with sufficiently heavy-tailed distributions, the median performs well. So, if one is willing to sacrifice affine equivariance, applied situations might arise where the usual median has practical advantages. In general, however, the OP estimator seems best for general use. It was found to be the most accurate alternative to the mean under normality, and it remains competitive under fairly extreme kurtosis.

As for probability coverage, when using method OP, Table 3 contains  $\hat{\alpha}$ , the estimated probability that the confidence region based on the Bonferroni method does not contain the population value when  $n = 20$  for  $p = 2$  and 4. For this portion of the study, 1,000 replications were used with  $B = 1,000$ . For asymmetric distributions, the actual value of the parameter was determined by taking the mean of 5,000 estimates based on a sample size of  $n = 100$ . Bradley (1978) argues that when testing at the .05 level, at a minimum the actual probability of a Type I error should be between .025 and .075. This criterion is satisfied in all cases except when  $p = 2$ ,  $(g, h) = (.5, 0)$  and  $\rho = 0$ , in which case  $\hat{\alpha} = .079$ . Increasing  $n$  to 30, the estimate equals .069.

Table 3: Estimated Type I Error Probabilities Using the OP Estimator,  $\alpha = .05, n = 20$ .

$g$	$h$	$\rho$	$p$	$\hat{\alpha}$
0.0	0.0	0.0	2	.071
0.0	0.0	0.7	2	.071
0.0	0.5	0.0	2	.040
0.0	0.5	0.7	2	.040
0.5	0.0	0.0	2	.079
0.5	0.0	0.7	2	.040
0.5	0.5	0.0	2	.056
0.5	0.5	0.7	2	.047
0.0	0.0	0.0	4	.065
0.0	0.0	0.7	4	.069
0.0	0.5	0.0	4	.040
0.0	0.5	0.7	4	.036
0.5	0.0	0.0	4	.063
0.5	0.0	0.7	4	.061
0.5	0.5	0.0	4	.044
0.5	0.5	0.7	4	.040

### Conclusion

A criticism of the OP estimator is that it is based on an outlier detection method that has a finite sample breakdown point of at most .25, because when using the interquartile range, the finite sample breakdown point is .25 for any projection. However, this would seem to suffice for many situations, and its efficiency is quite good compared to the mean even when sampling

from a very heavy-tailed distribution. If there are indications that more than 25% of the points are outliers, then one possibility is to use the variation of the OP estimator based on equation (3). The main point is that good efficiency is achieved under normality and a method for computing a confidence region was found that performs reasonably well in simulations.

#### References

- Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., & Tukey, J. W. (1972). *Robust estimates of location: survey and advances*. Princeton University Press, Princeton: NJ.
- Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, *31*, 144–152.
- Carling, K. (2000). Resistant outlier rules and the non-Gaussian case. *Computational Statistics & Data Analysis*, *33*, 249–258.
- Davies, P. L. (1987). Asymptotic behavior of S-estimators of multivariate location parameters and dispersion matrices. *Annals of Statistics*, *15*, 1269–1292.
- Devlin, S. J., Gnanadesikan, R., & Kettenring, J. R. (1981). Robust estimation of dispersion matrices and principal components. *Journal of the American Statistical Association*, *76*, 354–362.
- Donoho, D. L., & Gasko, M. (1992). Breakdown properties of location estimates based on halfspace depth and projected outlyingness. *Annals of Statistics*, *20*, 1803–1827.
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman and Hall.
- Fung, W. K. (1993). Unmasking outliers and leverage points: A confirmation. *Journal of the American Statistical Association*, *88*, 515–519.
- Golberg, K. M., & Iglewicz, B. (1992). Bivariate extensions of the boxplot. *Technometrics*, *34*, 307–320.
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. & Stahel, W. A. (1986). *Robust statistics*. New York: Wiley.
- Hoaglin, D. C. (1985). Summarizing shape numerically: The g-and-h distributions. In D. Hoaglin, F. Mosteller and J. Tukey (Eds.) *Exploring data tables, trends, and shapes*. (p. 461–515). New York: Wiley.
- Liu, R. Y., & Singh, K. (1997). Notions of limiting P values based on data depth and bootstrap. *Journal of the American Statistical Association*, *92*, 266–277.
- Liu, R. Y., Parelius, J. M., & Singh, K. (1999). Multivariate analysis by data depth: Descriptive statistics, graphics and inference. *Annals of Statistics*, *27*, 783–858.
- Masse, J. C., & Plante, J. F. (2003). A Monte Carlo study of the accuracy and robustness of ten bivariate location estimators. *Computational Statistics & Data Analysis*, *42*, 1–26.
- Peña, D., & Prieto, F. J. (2001). Multivariate outlier detection and robust covariance matrix estimation. *Technometrics*, *43*, 286–299.
- Poon, W. Y., Lew, S. F., & Poon, Y. S. (2000). A local influence approach to identifying multiple outliers. *British Journal of Mathematical and Statistical Psychology*, *53*, 255–273.
- Raine, A., Buchsbaum, M., & LaCasse, L. (1997). Brain abnormalities in murderers indicated by positron emission tomography. *Biological Psychiatry*, *42*, 495–508.
- Rocke, D. M. (1996). Robustness properties of S-estimators of multivariate location and shape in high dimension. *Annals of Statistics*, *24*, 1327–1345.
- Rocke, D. M., & Woodruff, D. L. (1996). Identification of outliers in multivariate data. *Journal of the American Statistical Association*, *91*, 1047–1061.
- Romanazzi, M. (1997). A schematic plot for bivariate data. *Student*, *2*, 149–158.
- Rosenberger, J. L., & Gasko, M. (1983). Comparing location estimators: Trimmed means, medians, and trimean. In D. Hoaglin, F. Mosteller and J. Tukey (Eds.) *Understanding robust and exploratory data analysis*. p. 297–336. New York: Wiley.
- Rousseeuw, P. J., & Leroy, A. M. (1987). *Robust regression & outlier detection*. New York: Wiley.

Rousseeuw, P. J., & Ruts, I. (1996). AS 307: Bivariate location depth. *Applied Statistics*, 45, 516–526.

Rousseeuw, P. J., & Ruts, I., & Tukey, J. W. (1999). The bagplot: A bivariate boxplot. *American Statistician*, 53, 382–387.

Rousseeuw, P. J., & Struyf, A. (1998). Computing location depth and regression depth in higher dimensions. *Statistics and Computing*, 8, 193–203.

Rousseeuw, P. J., & van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41, 212–223.

Rousseeuw, P. J., & van Zomeren, B. C. (1990). Unmasking multivariate outliers and leverage points (with discussion). *Journal of the American Statistical Association*, 85, 633–639.

Staudte, R. G., & Sheather, S. J. (1990). *Robust estimation and testing*. New York: Wiley.

Tukey, J. W. (1960). A survey of sampling from contaminated normal distributions. In I. Olkin, S. Ghurye, W. Hoeffding, W. Madow & H. Mann (Eds.), *Contributions to probability and statistics*. Stanford, CA: Stanford University Press.

Tukey, J. W. (1974). T6: Order statistics. *Mimeographed notes for Statistics 411*, Princeton University.

Tukey, J. W. (1975). Mathematics and the picturing of data. *Proceedings of the International Congress of Mathematicians*, 2, 523–531.

Tyler, D. E. (1991). Some issues in the robust estimation of multivariate location and scatter. In W. Stahel & S. Weisberg (Eds.), *Directions in robust statistics and diagnostics*, Part II, p. 327–336. New York: Springer-Verlag.

Viljoen, H., & Venter, J. H. (2002). Identifying multivariate discordant observations: a computer-intensive approach. *Computational Statistics & Data Analysis*, 40, 159–172.

Wilcox, R. R. (1997). *Introduction to robust estimation and hypothesis testing*. San Diego, CA: Academic Press.

Wilcox, R. R. (2003). *Applying contemporary statistical methods*. San Diego, CA: Academic Press.

Wilcox, R. R. (in press). Two-sample, bivariate hypothesis testing based on Tukey's depth. *Multivariate Behavioral Research*.

Zani, S., Riani, M., & Corbellini, A. (1998). Robust bivariate boxplots and multiple outlier detection. *Computational Statistics & Data Analysis*, 28, 257–270.

## Appendix

Following Liu and Singh (1993), Tukey's depth is defined as follows. Let  $F$  be a  $p$ -variate distribution. Tukey's depth at the point  $x$  is

$$TD(x; F) = \inf_H \{P(H) : H \text{ is a closed space containing } x\}.$$

The sample version is obtained by replacing  $F$  with the usual empirical distribution. More precisely, the sample version of  $TD(x; F)$  is the smallest proportion of  $X_i$  contained in any closed halfplane with boundary line through  $x$ . For  $p = 1$ ,  $TD(x; F) = \min\{F(x), 1 - F(x^-)\}$ .