C. CARONI

ADVANCES IN TESTING FOR MULTIVARIATE OUTLIERS

PLAN

ABSTRACT

This paper reviews difficulties arising in the use of Wilks' multivariate outlier test statistic and examines various modifications. These include the use of highly robust estimators and the development of a sequential version for testing for an unspecified number of outliers. When the covariance matrix has a special structure, a suitably adapted likelihood ratio test statistic offers a substantial increase in power over Wilks' test. Union-intersection testing can also offer improved power, but only in restricted circumstances. It is concluded that the sequential test offers the best way of identifying outliers in multivariate normal samples when a formal test of significance is required.

1. INTRODUCTION

An *outlier* in a sample of data is a point which appears substantially different from the rest. In a univariate sample, it will be an exceptionally large or small value. It is important in statistical analysis to identify points which are so extreme that it is doubtful that they come from the distribution that generated the other points, because this may highlight errors in the data or show the way to improve the statistical model. A substantial literature has thus developed on methods of detecting outliers (Hawkins, 1980a; Barnett & Lewis, 1994). The topic is quite closely related to the study of *influence*, since such extreme values will often have a large effect on parameter estimates.

However, relatively little has been written on the problem of detecting outliers in multivariate data, in contrast to the extensive literature that exists for the univariate case. The main reason for this is the greater difficulty —both analytical and computational— of the multivariate case. While many univariate methods are based on order statistics, since an outlier must be an extreme value, there is no simple concept of ordering multivariate data which would permit a simple extension of the univariate analysis. Furthermore, the multivariate case introduces a different kind of outlier —a point which is not extreme in any single dimension, but departs from the prevailing pattern of correlation between dimensions. Such methods as do exist are generally for the multivariate normal distribution. Again, difficulty is one reason for the restriction. There is a small literature on detecting outliers in specific non-normal multivariate distributions (Barnett & Lewis, 1994) but it is not very useful in practice.

Only Wilks' method (Wilks, 1963) for multivariate outlier detection could be said to be widely known and used. Despite its importance, it does have several deficiencies and limitations. The main purpose of this review is to discuss various extensions and modifications which may overcome some of these problems. The performance of the basic test statistic will be examined in Section 2 and the extensions will be considered thereafter. Section 3 discusses methods based on robust estimates, Section 4 describes the construction of a sequential test, Section 5 looks at tests for detecting outliers in specific correlation structures and, finally, Section 6 discusses procedures based on the union-intersection principle as an alternative to the likelihood ratio approach which leads to Wilks' statistic.

2. WILKS' TEST

Wilks' test can be derived by «two-stage maximum likelihood» (Barnett & Lewis, 1994) from a commonly-used model for generating outliers

$$\begin{array}{llll} H_0: & x_i \sim N_p(\mu, \Sigma) & & i = 1, 2, \ldots, n \\ H_1: & x_i \sim N_p(\mu, \Sigma) & & i \notin S \\ & x_j \sim N_p(\mu + a_j, \Sigma) & & j \in S \end{array}$$

where S is a set of k indices from {1,2,...,n}. The null hypothesis is that all n sample points are drawn from one multivariate normal distribution; under the alternative, n-k points are from this distribution but the remaining k are from distributions whose means have «slipped». The membership of S is not specified in advance although k is. That is, we say at this stage *how many* points should be tested as outliers, but not *which* ones.

For $k = 1$, and if the specific point x_j has been nominated as the potential outlier, the likelihood ratio test statistic for H_0 versus H_1 is easily shown to be

$$\Lambda_j = |A_j| / |A| \tag{1}$$

where A is the usual sample sum of squares and products matrix (SSP) and A_j is the SSP matrix of the reduced sample of n-1 points formed by omitting x_j. This is the first stage of the procedure. The second stage is to say that a reasonable choice of test statistic when no potential outlier can be nominated in advance is the minimum of (1) over all possible choices:

$$\min \Lambda_j = \min_j \ |A_j| / |A|$$

For $k > 1$, the corresponding procedure gives the test statistic

$$\min_S \ |A_S| / |A|$$

where S is any set of k indices from {1, 2,...,n}.

For $k = 1$, it can be shown that

$$\Lambda_j = 1 - \{n/(n-1)\} (x_j - \bar{x})' A^{-1} (x_j - \bar{x}) \ldots \tag{2}$$

and

$$\begin{aligned} \Lambda_j^{-1} &= 1 + \{(n-1)/n\} (x_j - \bar{x}_j)' A_j^{-1} (x_j - \bar{x}_j) \\ &= 1 + T_j^2 / (n-2) \end{aligned}$$

where $\bar{x}_j$ is the mean of the n–1 points remaining after x_j has been deleted and T_j^2 is Hotelling's T^2 statistic for testing the hypotheses

$$\begin{array}{llll} H_0: & x_i \sim N_p(\mu, \Sigma) & & i = 1, 2, \ldots, n \\ H_1: & x_i \sim N_p(\mu, \Sigma) & & i \neq j \\ & x_j \sim N_p(\mu + a_j, \Sigma) & & \end{array}$$

Thus this outlier testing problem (for a nominated potential outlier) is equivalent to a two-group comparison, a point that will be recalled in Section 6. Since T^2 is proportional to an F statistic, and using the well known relationship between the F and Beta distributions, it can be shown that

$$\Lambda_j \sim B((n-p-1)/2,\ p/2)$$

For the case k = 2, the distributional result for a nominated pair of potential outliers $\{x_i, x_j\}$ is

$$\Lambda_{ij}^{1/2} \sim B(n-p-2,\ p)$$

In general, Λ_S is distributed as the product of k independent Beta random variables.

Wilks' statistic is the minimum of several statistics which are not independent of each other. Its distribution is unknown even for the simplest case of k = 1. Wilks proposed using the above distributional results to give Bonferroni approximations to the percentage points required for testing.This is a common strategy in constructing outlier tests and it is generally true that the procedure gives a good test for the single outlier case but is liable to be very conservative for two or more outliers (Hawkins, 1980a). Table 1 (condensed from Caroni, 1990) gives selected simulation results showing the accuracy of the Bonferroni bounds, in the form of true exceedance probabilities (under the null) of tests conducted at the nominal significance level of 5%.

TABLE 1

Simulated exceedance probabilities of Bonferroni percentage points for Wilks' outlier test statistic, based on 40000 simulated samples for 2 outliers and 2000 samples for 3 and 4 outliers

		a = 1%			a = 5%		
		No. of outliers			No. of outliers		
n	p	2	3	4	2	3	4
10	2	.0082	.0075	–*	.0359	.0345	–
	3	.0085	.0090	–	.0374	.0335	–
	4	.0090	.0080	–	.0401	.0370	–
20	2	.0067	.0045	.0035	.0295	.0175	.0150
	3	.0070	.0060	–	.0297	.0150	–
	4	.0074	.0035	.0030	.0299	.0185	.0115
50	2	.0051	–	–	.0215	–	–
	3	.0058	–	–	.0244	–	–
	4	.0062	–	–	.0242	–	–

* Not computed

For two outliers, the true level tends to be only about half of the nominal level. For three and four outliers, the true size drops to as little as a quarter of the nominal. It is therefore evident that Wilks' test is not in fact a useful formal test statistic for testing for more than one outlier.

This presents an enormous practical problem, because —despite the construction of the statistic presented above— it is very rare for the potential number of outliers to be known. In practice, the analysis of the data must determine how many outliers there are and which points are outliers. As shown above, this cannot be done by testing for one, two, three outliers and so on, because after the first step we have inadequate approximations to the signifiacance levels. In addition, when we use an outlier test statistic to test for a specific number of outliers, we have to deal with the well known problems of masking and swamping which may tend to give rise to false conclusions. *Masking* occurs when there are more than k outliers and they are distributed in such a way that the k outlier test statistic is not significant at the chosen level, which could lead to the incorrect decision to stop testing at this stage. *Swamping* refers to the situation of having fewer than k outliers yet these being so extreme that they make the k outlier test statistic appear to be significant, leading to the decision to declare more than the correct number of outliers. In the next section, we examine a modification of Wilks' statistic to counter these problems.

As an alternative to successively testing for one, two, three and more outliers, another strategy is to test for one outlier then, if this test declares an outlier (rejects the null), omit the corresponding point and repeat the one-outlier test in the reduced sample of n-1 points. This procedure of reduction and re-testing is repeated until a stage when the null hypothesis is accepted. At each stage, the one-outlier test is used without any modification to allow for the fact that it is part of a larger procedure. However, since each stage is conditional on the previous one, it is obvious that the significance levels of the one-outlier test may not correspond to the significance level of the overall procedure. The point of the investigation is Section 4 is to find a way of conducting this procedure while retaining control of the overall level of significance.

3. ROBUST DETECTION OF OUTLIERS

The problems of masking and swamping arise in their different ways as a result of a common cause, namely that we are trying to detect outliers as points which are extreme in relation to the variability of the sample —yet, until we have eliminated outliers, we cannot have a good estimate of this variability. This contradiction can be avoided, or at least minimized, by applying robust estimation to the sample. In the case of Wilks' statistic, we may implement this in the form (2), inserting robust estimates which are unlikely to be much distorted by outliers.

This approach has been developed in particular by Rousseeuw who advocates his *minimum volume ellipsoid* (MVE) estimators (Rousseeuw, 1985). These have breakdown point 0.5; in other words, up to half of the sample points could be outliers without affecting the validity of the estimation of the mean and covariance matrix of the null distribution. The MVE is defined to be the ellipsoid which has minimum volume amongst all ellipsoids which cover at least half of the sample points. This can be thought of as a multivariate definition of the densest set of points in the sample, which is most unlikely to contain outliers and can therefore be the basis for uncontaminated estimation of the parameters of the null distribution.

In this approach, then, we would concentrate not on the ratios Λ_j but on the Mahalanobis distances: from (2), minimizing Λ_j is eqivalent to maximizing the distance. The population Mahalanobis distance of x_j from the mean is

$$(x_j-\mu)'\,\Sigma^{-1}\,(x_j-\mu)$$

and Wilks' statistic is equivalent to estimating this by inserting the usual estimators $\bar{x}$ and $S = A/(n-1)$ for μ and Σ respectively. Rousseeuw's approach is to replace them instead by the MVE estimators, giving robust distances. He then identifies which points appear to be outliers by comparing these robust distances against the percentage points of a X^2 distribution (Rousseeuw & Van Zomeren, 1990).

We observe that there are two principal drawbacks to this excellent idea. One is that the use of the X^2 reference distribution is only indicative. There is no proof of its applicability, so if in fact a formal test is required, this method does not give it. The second problem is that the computation involved is formidable. The method suggested by Rousseeuw & Van Zomeren (1990) actually only looks at a sample of the possible ellipsoids so may not find the true minimum, as well as having difficulty with almost collinear sets of points. A variation proposed by Hadi (1992) may well overcome the second drawback, but the lack of a formal test remains. Other adaptations of the method could include the use of different robust estimators, such as Campell's (1980) M-estimators, but would not avoid this difficulty in testing.

4. SEQUENTIAL APPLICATION OF WILKS' TEST STATISTIC

We now look at the procedure consisting of a sequence of applications of Wilks' one-outlier test statistic. In accordance with Hawkins (1980a), the preferred way to conduct this is given by the following steps.

(i) Choose a suitable value k, the maximum number of outliers which will be tested for in the data.

(ii) Find the k most extreme points of the sample as follows:

  (a) Use the one-outlier test statistic to define the most extreme point. Remove it from the sample (irrespective of its level of significance). This gives a reduced sample of n-1 points.

  (b) Repeat (a) in the reduced sample. Again remove the most extreme point and so on; repeat these steps until k points have been identified and removed.

(iii) Now test the point which was removed *last*, by performing the one outlier test in the sample of size n-k+1 which consists of this point and the n-k points which were not identified as extreme. If the result is significant at the chosen level, declare this *and all the other extreme points* to be outliers — that is, there are k outliers.

(iv) If the test is (iii) is not significant, put the next most extreme point back into the sample and test it using the one-outlier test in the sample of size n-k+2. If it is significant, declare this *and the k-2 more extreme points* to be outliers — that is, there are k-1 outliers.

(v) If it is not significant, put the next most extreme point back into the sample, test it, and so on.

The notable property of this procedure is that we cannot have masking, because we test for the higher number of outliers first, yet we cannot have swamping either, because at each stage the more extreme points are excluded from the computation of the test statistic. When applied to the particular case of Wilks' test, the procedure also has the enormous advantage that it only requires critical values for the one outlier test which, as we have seen, are the only accurate values we have for Wilks' test.

The above description leaves open the question of what significance level should be used at each step and what overall level is attained for the procedure as a whole. The levels chosen by Caroni & Prescott (1992), following Rosner's (1983) work in univariate cases, is as follows.

Let $D_1, D_2, \ldots, D_k$ be the values of the one-outlier test statistic in the steps defined above, with D_1 corresponding to the most extreme point. Let $\lambda_1, \lambda_2, \ldots, \lambda_k$ be the critical values for the corresponding tests. Then the following choice of $\{\lambda_i\}$ ensures that the probability of declaring *more than the true number* of outliers is equal to a selected level of significance *a whatever the true number of outliers is:*

$$\Pr\left\{\bigcup_{r=s+1}^{k}(D_r < \lambda_r | H_s)\right\} = \alpha \qquad s = 0, 1, \ldots, k-1,$$

where H_s is the hypothesis that there are s outliers. This definition was proposed by Hawkins (1980b), who criticized earlier definitions. It is equivalent to

$$\Pr\left\{\bigcap_{r=s+1}^{k}(D_r \geq \lambda_r | H_s)\right\} = 1-\alpha \qquad s = 0, 1, \ldots, k-1.$$

Rosner conjectured that this probability essentially depended on D_{s+1}, so that if

$$\Pr\left(D_{s+1} \geq \lambda_{s+1} | H_s\right) = 1-\alpha'$$

then α' is close to α. He found that this was indeed true in the univariate case he was studying, for sample sizes bigger than about 25 points. The investigation carried out by Caroni & Prescott (1992) using Wilks' statistic reached a similar conclusion for multivariate outlier detection. The practical meaning of this is that the procedure can be implemented by using the usual $100\alpha\%$ critical values for a one-outlier test at each step while still having the overall level of the entire procedure held at this same level. This avoids the need for special computing of new critical values which might otherwise be necessary.

Consequently, the sequential procedure avoids all the problems referred to earlier: masking, swamping, lack of accurate critical values for $k > 2$ and the unknown level of significance of simple repeated applications of the test statistic. It is worth emphasizing that its properties depend on the fact that the tests are conducted in the order described above, that is, starting from the least extreme.

5. EXTENSIONS OF WILKS' TEST FOR SPECIFIC CORRELATION STRUCTURES

Wilks' test is applicable to a simple random sample from the distribution $N_p(\mu, \Sigma)$ where Σ is any non-singular covariance matrix. This is a complete contrast to many situations in multivariate analysis where we assume or investigate specific covariance structures –for example, canonical correlations analysis or factor analysis. It is worth considering the development of outlier detection techniques which are applicable to particular structures.

One important special structure for Σ is the equicorrelation matrix

$$\Sigma = \sigma \{ \rho J + (1-\rho) I \}$$

where I is the identity matrix and J is the pxp matrix whose elements are all unity. By applying the two-stage maximum likelihood method, the following statistic for testing for one outlier under the slippage-of-the-mean model can be derived

$$\min_j \left(\frac{\hat{\sigma}_j^2}{\hat{\sigma}^2} \right)^p \frac{|\hat{R}_j|}{|\hat{R}|}$$

where $\hat{R}$ is the maximum likelihood estimator of the correlation matrix, with off-diagonal elements

$$\hat{\rho} = \left\{ \sum_{i \neq j} \sum a_{ij} / n \right\} / \left\{ p(p-1)\hat{\sigma}^2 \right\}$$

where a_{ij} are the elements of A and $\hat{R}_j$ is the equivalent estimator after deleting x_j from the sample (Caroni & Prescott, 1991). As with Wilks' test, its form is the ratio of two determinants, one computed from the full sample and the other from the reduced sample.

Since the distribution of the test statistic for a nominated potential outlier is unknown, it is not possible to derive Bonferroni bounds for testing. Hence the power of this test in comparison to Wilks' test has been investigated using simulated critical values. A selection of results is shown in Table 2. As expected, the new statistic, which incorporates the information on the structure of the matrix, is more powerful than Wilks' ordinary test, which does not. The gain is very large for small and medium sample sizes, up to about n = 30.

Another special structure for Σ is the block structure (e.g., canonical correlation analysis)

$$\Sigma = \begin{pmatrix} \Sigma_{11} & 0 \\ 0 & \Sigma_{22} \end{pmatrix}$$

where there are two uncorrelated sets of variables. The test statistic in this case is (Caroni & Prescott, 1991)

$$\min_j \frac{|A_{1j}| \cdot |A_{2j}|}{|A_1||A_2|}, \qquad j = 1, \ldots, n$$

where the subscripts 1 and 2 denote the two subsets of variables. Since, for a specified potential outlier, this statistic is distributed as the product of two independent Betas, it is possible to derive Bonferroni critical values, at least in particular cases. As in the previous case, important gains in power can be obtained by using this test in preference to Wilks' test.

Other special structures, such as the factor analysis model or the correlations arising in autoregressive processes, have not yet been investigated from the point of view of outlier detection, but it is to be expected that a test statistic specifically adapted to the model will always be worth using in preference to the general statistic.

TABLE 2

Comparison between simulated powers of outlier test statistic for equicorrelation model (EC) and Wilks' outlier test (W). Entries for each combination of n and p are based on 6000 simulated samples. Percentage points for EC and W were simulated from 14000 samples at each combination of n and p.

		$\alpha = 1\%$		$\alpha = 5\%$	
n	*p*	*EC*	*W*	*EC*	*W*
10	2	.1417	.0942	.3438	.2915
	3	.1137	.0503	.2818	.1865
	4	.0885	.0268	.2482	.1148
	5	.0747	.0165	.2228	.0837
20	2	.2580	.2258	.4673	.4465
	3	.1953	.1658	.4075	.3450
	4	.1682	.1060	.3553	.2712
	5	.1238	.0632	.2845	.1950
30	2	.2942	.2890	.5008	.4850
	3	.2272	.2112	.4215	.3880
	4	.1863	.1403	.3630	.3163
	5	.1635	.1017	.3298	.2590
50	2	.3202	.3070	.5157	.5090
	3	.2498	.2388	.4442	.4338
	4	.2215	.1832	.3835	.3568
	5	.1785	.1417	.3417	.2985

6. THE UNION-INTERSECTION APPROACH TO MULTIVARIATE OUTLIER DETECTION

After likelihood ratio, the union-intersection method is the second general approach to hypothesis testing in the multivariate case. It is natural to examine what it offers for outlier detection as an alternative to the likelihood ratio approach which yields Wilks' statistic.

The appropriate test statistic can be found by looking at the outlier detection problem as a comparison between groups, as mentioned in Section 2 above. More specifically, the k-outlier problem can be represented as a one-way MANOVA between k+1 groups. One group consists of the n-k «good» points (that is, points from the null distribution) and the remaining k groups each consist of a single point (the k potential outliers in the set S). Within this framework, the matrix A_s is the same as the within-group SSP matrix usually denoted by W (since only the group with more than one point contributes) while the matrix A is the total SSP, T = W+B where B is the between-group SSP matrix. The likelihood ratio test statistic can be expressed as

$$\begin{aligned}\Lambda_s &= |A_s|/|A| \\ &= |W|/|W+B| \\ &= \prod_{j=1}^{p}(1+\lambda_j)^{-1}\end{aligned}$$

where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$ are the eigenvalues of $W^{-1}B$. The union-intersection test statistic for the same situation is (e.g., Mardia, Kent & Bibby, 1979)

$$\vartheta = \lambda_1/(1+\lambda_1)$$

which is the largest eigenvalue of $(W+B)^{-1}B$. This applies for a specified set of potential outliers, so we implement a «two-stage union-intersection» method and maximize the value of ϑ over all possible choices of points.

Since λ_1 is the only non-zero eigenvalue when k = 1, the union-intersection and likelihood ratio tests coincide for the one-outlier case. To compare the merits of the two approaches, it is necessary to look at tests for at least two outliers. We note (Section 2) that the Bonferroni critical values for Wilks' two-outlier test are inadequately accurate, and the same has been confirmed for the above union-intersection test statistic (Bonferroni bounds can be computed using Khatri's series expansion of the distribution function of the greatest-root statistic; see Venables, 1975). Therefore, a meaningful comparison between the powers of the two tests can only be carried out by first obtaining simulated critical values so that the true sizes of the tests are known. The results of this comparison (Caroni & Prescott, 1995) accord with what would be predicted from knowledge of similar comparisons of the rival test statistics for MANOVA. In particular, if group means are almost collinear, so that one eigenvalue of B is much bigger than the rest, the union-intersection test is more powerful than likelihood ratio. On the other hand, the likelihood ratio test is better in other situations, since it uses all the information and not just that in the first component. Figure 1 illustrates the equivalent result for the two-outlier test, showing the relative powers of the two tests for detecting two outliers in samples of size 30 in two and five dimensions. The two outliers were generated by adding slippages to randomly selected members of the samples. One slippage was fixed while the other was of the same magnitude, but varied in direction relative to the other from 0° to 180° in the study.

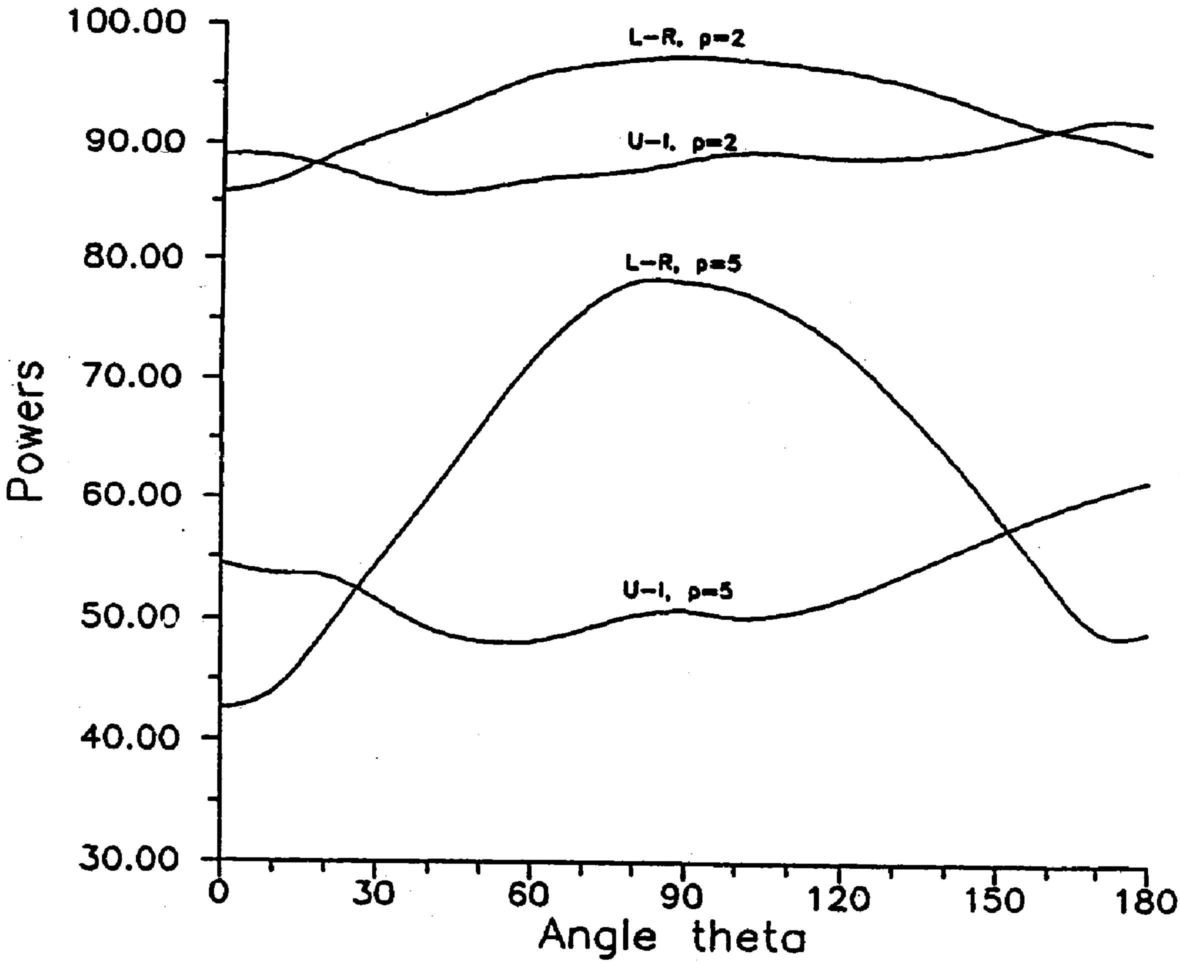

Figure 1. Powers of union-intersection (U-I) and Wilks' (L-R) tests with varying angle θ between slippage vectors, at 1% level of significance.

As predicted, the union-intersection test is the more powerful when the angle between the two vectors is close to 0° or 180° —that is, when the slippages are almost collinear. However, its advantage over the likelihood ratio test is relatively small. On the other hand, there is a much wider range over which the likelihood test is the more powerful and its advantage over the union-intersection test can be very big. For this reason, the conclusion is that the union-intersection approach cannot be recommended in preference to tests based on likelihood ratio, unless it is safe to assume the collinear structure. What this assumption actually requires, in the slippage-of-the-mean model, is that outliers occur only in a particular direction, although with variable magnitude. This is not likely to be a mechanism that arises very often in real life.

7. DISCUSSION

It is no exaggeration to say that Wilks' test is virtually all that is available for outlier detection in the multivariate case. Apart from its direct use, the test is the basis of several other methods, such as a graphical implementation (Bacon-Shone & Fung, 1987) and the successive version discussed in Section 4 above, while the logic of its likelihood ratio derivation leads to the statistics of similar form in Section 5 above. Even Rousseeuw's method, which at first sight looks rather different, is only the same test with robust estimators taking the place of the usual ones. Ideas with an essentially different approach from Wilks' method include the gap test proposed by Rohlf (1975) and the union-intersection analysis given in Section 6 above. However, the former does not provide a usable formal test of significance (Caroni & Prescott, 1995), while the latter appears to be useful only in special circumstances.

In this review, there has been an emphasis on formal tests of significance. If one approaches outlier detection as part of an exploration of a data set, it may be felt that formal tests are unnecessary or, at least, that rough ones are adequate. In this case, Rousseeuw's method may be recommended, although preferably with improvements to the computational method. Another approach which seems to be extremely promising is the robust principal components analysis (RPCA) (Campbell, 1980; Matthews, 1984), which provides the usual insight into data structure offered by a principal components analysis and additionally suggests outliers through the pattern of weights attached to individual points in the estimation.

On the other hand, it is the author's claim that it is essential to have formal tests available. One reason is that outlier detection should in some circumstances be a procedure applied routinely in data processing –for example, to batches of laboratory measurements– and then the objectivity of the formal test is important. Another reason is that it is quite probably true that people who look for outliers tend to believe that they have found them, then omit these points from their data and thereby end up with an excessively homogeneous sample which fits the preconceived model very well. The objectivity of the formal test guards against this over-enthusiastic rejection of extreme points which are not in fact outliers (Hawkins, 1980a).

When a formal test is required the recommended test is the sequential version of Wilks' test or, if a specific correlation structure can be assumed, an equivalent likelihood ratio test for data with that structure. Some informal method such as RPCA should probably be used as well, but only to aid understanding of the data and not as the basis for a decision. These recommendations apply, of course, to data for which the multivariate normal distribution can be assumed, after appropriate transformation if necessary. If the data follow another distribution, the claimed significance levels of the tests are meaningless, but neither are the informal methods reliable. As remarked in the introdution, Barnett has made a start on the development of outlier detection methods for non-normal multivariate distributions, but his «attempt to awaken interest in the topic» (Barnett, 1979) has not yet led to much response from other investigators.

REFERENCES

Bacon-Shone J. & Fung W.K. (1987), «A new graphical method for detecting single and multiple outliers in univariate and multivariate data», *Appl. Statist.*, 36, 153-162.

Barnett V. (1979), «Some outlier tests for multivariate samples», *South African Statist. J.*, 13, 29-52.

Barnett V. & Lewis T. (1994), *Outliers in Statistical Data*, (3rd ed), Wiley, Chi- chester.

Campbell N.A. (1980), «Robust procedures in multivariate analysis. I. Robust covariance estimation», *Appl. Statist.*, 29, 231-237.

Caroni C. (1990), *Multivariate Statistical Outliers*, Unpublished Ph.D. thesis, University of Southampton, U.K.

Caroni C. & Prescott P. (1991), «Multivariate outlier tests with structured covariance matrices», *J. Statist. Comput. Simul.*, 38, 165-179.

Caroni C. & Prescott P. (1992), «Sequential application of Wilks's multivariate outlier test», *Appl. Statist.*, 41, 355-364.

Caroni C. & Prescott P. (1995), «Union-intersection testing for outliers in multivariate normal data», *J. Statist. Comput. Simul.*, 51, 185-196.

Caroni C. & Prescott P. (1995), «On Rohlf's method for the detection of outliers in multivariate data», *J. Multivar. Anal.*, 52, 295-307.

Hadi A.S. (1992), «Identifying multiple outliers in multivariate data», *J. R. Statistic. Soc.*, B, 54, 761-771.

Hawkins D.M. (1980a), *Identification of Outliers*, Chapman & Hall, London.

Hawkins D.M. (1980b), «Critical values for identifying outliers», *Appl. Statist.*, 29, 95-96.

Mardia K.V., Kent J.T. & Bibby J.M. (1979), *Multivariate Analysis*, Academic Press, London.

Matthews J.N.S. (1984), «Robust methods in the assessment of multivariate normality», *Appl. Statist.*, 33, 272-277.

Rohlf F.J. (1975), «Generalization of the gap test for the detection of multivariate outliers», *Biometrics*, 31, 93-101.

Rosner B. (1983), «Percentage points for a generalized ESD many-outlier procedure», *Technometrics*, 25, 165-172.

Rousseeuw P. J. (1985), «Multivariate estimation with high breakdown point», In: *Mathematical Statistics and Applications*, (eds. W. Grossman, G. Pflug, I. Vincze and W. Wertz), Vol. B, pp. 283-297, Reidel, Dordrecht.

Rousseeuw P. J. & Van Zomeren B.C. (1990), «Unmasking multivariate outliers and leverage points (with Comments)», *J. Am. Statist. Ass.*, 85, 633-651.

Venables W.N. (1975), «Algorithm AS77: Null distribution of the largest root statistic», *Appl. Statist.*, 24, 458-465.

Wilks S.S. (1963), «Multivariate statistical outliers», *Sankhya*, A, 25, 407-426.