Available online at www.sciencedirect.com
SCIENCE DIRECT®

Computational Statistics & Data Analysis 49 (2005) 863–874

COMPUTATIONAL STATISTICS & DATA ANALYSIS

www.elsevier.com/locate/csda

# A robust testing procedure for the equality of covariance matrices

Shagufta Aslam[a,*], David M. Rocke[b]

[a]*Center of Image Processing and Integrated Computing, University of California, Davis, 2343 Academic Surge Building, Davis 95616, USA*

[b]*Department of Applied Science, University of California, Davis, 2343 Academic Surge Building, Davis 95616, USA*

Received 28 April 2003; received in revised form 11 June 2004; accepted 12 June 2004
Available online 6 July 2004

**Abstract**

In classical statistics the likelihood ratio statistic used in testing hypotheses about covariance matrices does not have a closed form distribution, but asymptotically under strong normality assumptions is a function of the $\chi^2$-distribution. This distributional approximation totally fails if the normality assumption is not completely met. In this paper we will present multivariate robust testing procedures for the scatter matrix $\Sigma$ using S-estimates. We modify the classical likelihood ratio test (LRT) into a robust LRT by substituting the robust estimates in the formula in place of classical estimates. A nonlinear formula is also suggested to approximate the degrees of freedom for the approximated Wishart distribution proposed for S-estimates of the shape matrix $\Sigma$. We present simulation results to compare the validity and the efficiency of the robust likelihood test to the classical likelihood test.


*Keywords:* S-estimate; Likelihood ratio test

## 1. Introduction

The estimation of covariance matrices may be called the key step to multivariate analysis. Robust estimators of these matrices open the door to the robustification of classical normal-theory multivariate procedures. The development of testing procedures for the

* Corresponding author. Tel.: +1-530-7647826; fax: +1-530-7528894.
*E-mail addresses:* saslam@ucdavis.edu, saslam@wald.ucdavis.edu (S. Aslam), dmrocke@ucdavis.edu (D.M. Rocke).

0167-9473/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2004.06.009

multivariate shape parameter is one of the most difficult problems in robust statistics. In classical theory, the likelihood ratio test (LRT) is the most commonly used test in the multivariate case because of its general asymptotic chi-square property. The LRT is based on the sample covariance $S$ which is described by the Wishart distribution $(W_p(\Sigma, n-1))$, provided that the samples are from $N_p(\mu, \Sigma)$. This test is not at all robust. The type-I error of the LRT is very sensitive to distributional assumptions. One way to decrease the sensitivity of the LRT to distributional assumptions is to replace the sample covariance with some high quality robust estimate of the covariance matrix for multivariate normals.

S-estimators are highly resistant to outliers and give essentially the same values as the usual analysis when there are no outliers. One of the plausible distributions one could think of to describe the S-estimate of shape matrix of multivariate normal is the Wishart distribution. The analogous behavior of S-estimators to the sample covariance $S$ with reference to its distribution and asymptotic unbiasedness led us to think about the development of a robust likelihood ratio test (RLRT) based on the S-estimate. In this paper we develop a RLRT for the following test on covariance matrices of a multivariate normal:

$$\mathrm{H}_0 : \Sigma_1 = \Sigma_2 = \cdots = \Sigma_r .$$

### 1.1. Definition: LRT

If the distribution of the random sample $X = (x_1, x_2, \ldots, x_n)'$ depends upon a parameter vector $\theta$, and if $\mathrm{H}_0 : \theta \in \Omega_0$ and $\mathrm{H}_a : \theta \in \Omega_1$ are any two hypotheses, then the likelihood ratio (LR) statistic for testing $\mathrm{H}_0$ against $\mathrm{H}_a$ is defined as:

$$\lambda(x) = L_0^*/L_1^*,$$

where $L_i^*$ is the largest value which the likelihood function takes in region $\Omega_i$, $i = 0, 1$.

The LRT of size $\alpha$ for testing $\mathrm{H}_0$ against $\mathrm{H}_a$ has as its rejection region

$$R = \{x | \lambda(x) < c\}, \tag{1.1}$$

where $c$ is determined so that

$$\sup_{\theta \in \Omega_0} P_\theta\{x \in R\} = \alpha. \tag{1.2}$$

For the hypotheses we are interested in, the distribution of $\lambda$ does not in fact depend on the particular value of $\theta \in \Omega_0$, so the supremum is unnecessary.

In general the exact density functions of LR statistics in multivariate analysis are so complicated that they appear to be of limited usefulness except for some special cases for which the distribution is quite tractable (Muirhead, 1982). The LRT has a very important asymptotic property given in following result:

> If $\Omega_1$ is a region in $R^q$ and if $\Omega_0$ is an $r$-dimensional subregion of $\Omega_1$, then under suitable regularity conditions for each $\theta \in \Omega_0$, $-2\log\lambda$ has an asymptotic $\chi^2_{q-r}$ distribution as $n \to \infty$.

LR statistics are based on maximum likelihood estimators. The MLE of $\Sigma$ and $\mu$ for $N_p(\mu, \Sigma)$ are the sample covariance matrix $S$ and sample mean $\bar{X}$. These estimates are very sensitive to outliers and long tail distributions. This makes the LR statistic also non-robust to outliers besides its sensitivity to the distribution. A simple way to robustify the statistic is to replace the nonrobust estimate with a high quality robust estimate. A basic criterion for a good robust estimator is its high breakdown and equally good behavior with uncontaminated data sets.

## 2. S-estimates

S-estimators originated in the regression context (Rousseeuw and Yohai, 1984) as a constraint optimization problem. Later, they were applied to the multivariate scale and location estimation problem (Davies, 1987). Lopuhaä showed that an S-estimate of location and scale is a type of constrained M-estimate (Lopuhaä, 1989). An S-estimate of multivariate location and shape is defined as follows:

### *2.1. Definition (Rousseeuw and Yohai, 1984)*

Let $\rho : \Re^+ \to \Re^+$ be a twice continuously differentiable, symmetric, nondecreasing function which has $\rho(0) = 0$ and is constant at $\rho(x) = \rho(c)$ for all $x \geqslant c$. Given a data set of $n$ points in $\Re^p$, let the S-estimator, $(\tilde{\mu}, \tilde{\Sigma})$, be defined by minimizing $|\tilde{\Sigma}|$ subject to

$$n^{-1} \sum_i \rho(d_i) = b_0,$$

where

$$d_i^2 = (x_i - \mu)^t \Sigma^{-1} (x_i - \mu)$$

and $b_0 = E(\rho(d))$ with $d \sim N(0, 1)$. Choice of $c$ depends on the desired breakdown point for the estimate.

We chose to use the translated biweight (t-biweight) of Rocke (1996) to provide the lowest sensitivity to outliers for a given breakdown point. Following is the t-biweight function:

*Translated-Biweight* (a,b) (Rocke, 1996):

$$\rho_{tb} = \begin{cases} \dfrac{d^2}{2} & d \leqslant a, \\ \left(\dfrac{a^2}{2} - \dfrac{a^2(a^4 - 5a^2b^2 + 15b^4)}{30b^4}\right. & \\ +d^2\left(\dfrac{1}{2} + \dfrac{a^4}{2b^4} - \dfrac{a^2}{b^2}\right) + d^3\left(\dfrac{4a}{3b^2} - \dfrac{4a^3}{3b^4}\right) & \\ \left. +d^4\left(\dfrac{3a^2}{2b^4} - \dfrac{1}{2b^2}\right) - \dfrac{d^5 4a}{5b^4} + \dfrac{d^6}{6b^4}\right) & a \leqslant d \leqslant a + b, \\ \dfrac{a^2}{2} + \dfrac{b(5b + 16a)}{30} & d \geqslant a + b. \end{cases}$$

### 2.2. Definitions: asymptotic rejection point (Rocke, 1996)

Consider a redescending M- or S-estimator, in which $c_0 = \inf\{d_0 \parallel \omega(d) = 0, \forall d \geqslant d_0\}$, where $\omega = \psi(d)/d$ and $\psi = \partial\rho(d)/\partial(x)$. The asymptotic rejection probability (ARP) of this estimator is then defined as the probability in a large sample under a reference distribution (usually multivariate normal) that the Mahalanobis distance exceeds $c_0$. If the estimator is normed, the ARP is $1 - F_{\chi^2(p)}(c_0^2)$, where $F$ is

$$F(x) = Pr\{(X-\mu)^{\mathrm{T}}\Sigma^{-1}(X-\mu) \leqslant x\} \tag{2.1}$$

and $p$ is the dimension of data matrix $X$.

The ARP gives the percentage of data points that would be given zero weight if in fact the data were uncontaminated (i.e. distributed multivariate normal). It is clear that we can choose $c_0$ to give any value for the ARP, but manipulating the value of the ARP will also change the breakdown of the estimator.

Though we are not constrained by only one parameter, not every combination of breakdown and ARP is possible with the t-biweight. Our programs are set to have the maximum breakdown, and the user is able to choose the ARP. If the user chooses an ARP that is too large, the parameter $b$ is reduced, and as $b \to 0$ the limit of the t-biweight is the Winsorized squares estimator. If the user chooses an ARP that is too small, the parameter $a$ is reduced, and as $a \to 0$, the limit of the translated biweight is the biweight estimator. Using the t-biweight, the two parameters $a$ and $b$ can be chosen to give the desired breakdown and ARP subject to the estimator changing in the limit. In other words, given a particular breakdown point, the ARP of the t-biweight cannot be larger than that of the least Winsorized squares estimator or smaller than that of the biweight estimator (Rocke, 1996).

With the t-biweight function beyond $a + b$, data point will be given zero weight in the estimating equations. The value $a$ will be calculated to give the correct breakdown. In this work we set $r = (n-p)/2n$, as this allows for the breakdown possible for S-estimators.

## 3. Distribution of S-estimates of $\Sigma$ in $N_p(\mu, \Sigma)$

Lopuhaä (1989) and Davies (1987) derived the asymptotic distributions of S-estimators of location and the shape matrix for multivariate normal. The main objective of this paper is to develop the testing procedures for the scatter parameter $\Sigma$ using its S-estimate for a data set of given dimension $n \times p$. This requires some knowledge about the distributional properties of this estimate. In the univariate case the simulation results showed that S-estimates of the dispersion parameter for $N(0, \sigma^2)$ can be approximated by a scale adjusted $\chi^2$ distribution with specified degrees of freedom (Aslam and Rocke, 2003). One cannot immediately generalize this for the higher dimensions. For small samples S-estimates behave equally well as the classical estimate when the underlying distribution is uncontaminated and the asymptotic properties of S-estimate for $\Sigma$ are comparable with the asymptotic properties of its classical estimate. This led us to think that similar to the classical sample covariance matrix a scale adjusted Wishart distribution is indeed the most plausible candidate to be considered for the approximate distribution of an S-estimate of such type of parameter when the underlying distribution is multivariate normal. That is, for a sample data matrix

$X_{n\times p}$ from $N_p(0, \Sigma)$, $mS_{tb}$ may approximately follow $cW_p(\Sigma, m)$, where $c$ is a constant satisfying

$$E(S_{tb}) = c\Sigma \tag{3.1}$$

(which holds for some $c$ because $S_{tb}$ is an affine equivariant shape estimate of $\Sigma$ (Tyler, 1983)), and $m$ is the degrees of freedom. Using Mardia et al. (1997) the relationship between the chi-square and Wishart distributions implies.

$$mc^{-1}s_{ii} \sim \chi^2_m \sigma^2_{ii}, \tag{3.2}$$

where $s_{ii}$ and $\sigma_{ii}$ are the diagonal elements of $S_{tb}$ and $\Sigma$, respectively. Since the estimators are affine equivariant, we can assume that the data came from $N_p(0, I)$ and therefore $\sigma_{ii} = 1$ and the diagonal elements are identically distributed (Grübel and Rocke, 1990). $S_{tb}$ is an affine equivariant estimator, so the distribution of it will have the same Wishart degrees of freedom for any $\Sigma$. Therefore, the estimates of $c$ and $m$ in the $N_p(\mu, \Sigma)$ case will be the same as the estimates of $c$ and $m$ in the $N_p(0, I)$ case.

### 3.1. Estimation of constants m and c

Now to approximate a distribution for $S_{tb}$ for small samples the constants $c$ and $m$ have to be estimated correctly. Using Hardin's (2000) method of moments identification, the estimates of $m$ and $c$ are determined by first two moments

$$\hat{m} = 2/\widehat{CV}, \tag{3.3}$$

$$\hat{c} = 1/p \sum_{i=1}^{p} s_{ii}, \tag{3.4}$$

where $\widehat{CV}$ is the estimated coefficient of variation of the diagonal elements of an S-estimate of the shape matrix. Since S-estimates of the covariance matrix are scaled to be consistent for the population covariance (under the multivariate normality assumption) (Davies, 1987), consistency gives us asymptotic unbiasedness, so asymptotically $\hat{c} = 1$. Thus, only the second moment is needed to find the desired degrees of freedom. Therefore,

$$\hat{m} = 2/\widehat{Var(s_{ii})}, \tag{3.5}$$

where $\widehat{Var(s_{ii})}$ is the estimated variance of the diagonal elements of an S-estimate of the covariance matrix $\Sigma$.

We can estimate the variance in two ways: through simulation and through an asymptotic formula (Lopuhaä, 1989; Davies, 1987). Since the diagonal elements are identically distributed and uncorrelated, we can simulate $N = 1000$ copies of the $p \times p$ S-estimate of shape matrix from the $n$ data points in each sample and then estimate $\hat{m}$ by using the variance of the $Np$ diagonal elements. Lopuhaä (1989) derived the value of variance of the diagonal elements under standard multivariate normality as

$$Var[s_{ii}] = \frac{2\sigma_1 + \sigma_2}{n}, \tag{3.6}$$

where

$$\sigma_1 = \frac{p(p+2)E_{0,I}[\psi^2(\| X \|) \| X\|^2]}{E_{0,I}[\psi'(\| X \|) \| X\|^2 + (p+1)\psi(\| X \|) \| X \|]}, \tag{3.7}$$

$$\sigma_2 = -2/p\sigma_1 + \frac{4E_{0,I}[(\rho(\| X \|) - b_0)^2]^2}{E_{0,I}[\psi(\| X \|) \| X \|]}, \tag{3.8}$$

$$b_0 = E_{0,I}[\rho \| X \|], \tag{3.9}$$

where $E_{0,I}$ is the expectation under standard multivariate normality, and $\| X \|$ is the norm of a vector $X$ in $R^p$.

### 3.2. Empirical degrees of freedom for S-estimators

It will make no difference in the results if we use the asymptotic degrees of freedom instead of simulated ones to approximate the Wishart distribution for $S_{tb}$ provided that the sample size is very large. But for small samples the actual and asymptotic degrees of freedom are quite far apart. The simulation is not an easy job particularly in higher dimensions. Therefore, it is good to have a handy mathematical model to approximate the degrees of freedom in real life situation. The simulation study showed that the degrees of freedom estimated by the simulation gets closer and closer to the ones computed by asymptotic variances but never exceed it as $n$ and $p$ get large. These observations from the simulated and asymptotic values of the degrees of freedom led us to develop an empirical relationship of the actual degrees of freedom with the asymptotic variance, $n$ (sample size) and $p$ (dimension of the data). The following non-linear models fits well:

$$\frac{m_{\text{simulated}}}{m_{\text{asymptotic}}} = 1 \pm \alpha \exp(\beta n + \gamma p). \tag{3.10}$$

The values of $\alpha$, $\beta$ and $\gamma$ were estimated by fitting the nonlinear regression curve to the simulated and asymptotic degrees of freedoms for different values of $n$ and $p$. The non-linear equation (3.10) is first linearized by using Taylor's expansion of $\exp(x)$ and can be written as

$$\begin{aligned}\frac{m_{\text{simulated}}}{m_{\text{asymptotic}}} &\approx 1 \pm \alpha(1 + \beta n + \gamma p)\\ &\approx 1 \pm (\alpha + \alpha\beta n + \alpha\gamma p).\end{aligned}$$

The initial estimates of $\alpha$, $\beta$ and $\gamma$ were obtained by fitting the linear regression to the data and using the formulas

$$\text{intercept} = 1 \pm \hat{\alpha}_{\text{start}},$$

$$\hat{\beta}_{\text{start}} = \frac{\text{coefficient on } n}{\hat{\alpha}_{\text{start}}},$$

$$\hat{\gamma}_{\text{start}} = \frac{\text{coefficient on } p}{\hat{\alpha}_{\text{start}}}.$$

Using the nonlinear regression function from S-plus we got the following fit to the simulated data with 3% residual standard error:

$$\frac{m_{\text{simulated}}}{m_{\text{asymptotic}}} \approx (1 - 0.449\ \exp(-0.00078n - 0.17128p)).$$

The above estimated non-linear regression is based on the values of $n$ ranges between 10 and 500 for values of $p$ ranges between 2 and 10. It is a fact that simulation is time consuming especially in high dimensions. Once a computer code is available for computing the asymptotic degrees of freedom on any computer package (S-plus, Fortran, Mathematica), one could get an approximate degrees of freedom for given values of $n$ and $p$. For simplicity of notation the author used in rest of the paper $v$ instead of $m_{\text{simulated}}$ to get the empirical number of degrees of freedom.

$$v \approx m_{\text{asymptotic}}(1 - 0.449\ \exp(-0.00078n - 0.17128p)). \tag{3.11}$$

## 4. RLRT

One of the reasons for the lack of robustness of the LRT is related to the poor robustness property of the sample covariance matrix as an estimate of the scatter matrix. This suggests that replacing the sample estimate $S$ by a robust estimate of the shape matrix in the LR statistic will produce a test with less sensitivity to the distributional assumptions.

### 4.1. Testing equality of covariance matrices

In this section we consider testing the null hypothesis that the covariance matrices of $r$ multivariate normal distributions are equal, given independent samples from these $r$ populations. Let $X_{i1}, \ldots, X_{in_i}$ be independent $N_p(\mu_i, \Sigma_i)$ random vectors $i = 1, 2, \ldots, r$ and consider the null hypothesis

$$\mathrm{H}_0 : \Sigma_1 = \cdots = \Sigma_r$$

against the alternative $\mathrm{H}_a$ which says that $\mathrm{H}_0$ is not true. Under $\mathrm{H}_0$ the common covariance matrix is unspecified. The assumption of equal covariance matrices is important in multivariate analysis of variance and discriminant analysis.

The classical modified LR statistic, suggested by Bartlett (1937) for $r = 2$, is defined as

$$\Lambda^{\star} = \frac{(|S_1|)^{n_1/2}(|S_2|)^{n_2/2}}{(|\frac{n_1S_1+n_2S_2}{n_1+n_2}|)^{(n_1+n_2)/2}}. \tag{4.1}$$

This LR statistic was obtained by working with the likelihood function of $\Sigma_1$ and $\Sigma_2$ specified by the joint marginal density function of $n_1S_1$ and $n_2S_2$ ( a product of Wishart distribution), rather than the likelihood function specified by the original normally distributed variables. The modified LRT then rejects $\mathrm{H}_0 : \Sigma_1 = \Sigma_2$ for small values of $\Lambda^{\star}$. The unbiasedness of this test was established in the univariate case $p = 1$ by Pitman (1939). For $r = 2$, this test is a uniformly most powerful unbiased test. The unbiasedness for general $p$ and $r$ was proved by Perlman (1980).

The asymptotic distribution for $-2\rho \log \Lambda^{\star}$, can be expanded for large $M = \rho n$ (Muirhead, 1982) as

$$P(-2\rho \log \Lambda^{\star} \leqslant x) = P(\chi_f^2 \leqslant x) + \gamma/M^2[P(\chi_{f+4}^2 \leqslant x) - P(\chi_f^2 \leqslant x)] + \mathrm{O}(M^{-3}), \tag{4.2}$$

where

$$n = n_1 + n_2,$$

$$f = p(p+1)(r-1)/2,$$

$$\rho = 1 - \frac{(2p^2+3p-1)}{6(p+1)(r-1)n}\left(\left(\sum_{i=1}^{r} 1/k_i\right) - 1\right),$$

$$k_i = \frac{n_i}{\sum_{i=1}^{r} n_i},$$

$$\gamma = M^2 \omega_2,$$

$$\omega_2 = \frac{p(p+1)}{48(n\rho)^2}\left[(p-1)(p-2)\left(\left(\sum_{i=1}^{r} 1/k_i^2\right) - 1\right) - 6(r-1)[n(1-\rho)]^2\right].$$

Earlier we mentioned that the Wishart distribution is the most plausible distribution to describe S-estimates of the shape matrix of the multivariate normal distribution. This conjecture along with the fact that the derivation of the classical LR statistic $\Lambda^*$ involves the product of two Wishart distributions associated with the sample covariance matrix motivated us to derive a robust modified LR statistic by replacing sample covariances $S_i$ with S-estimates $S_{tb}^{(i)}$ and $n_i$ with the corresponding degrees of freedom $v_i$, i.e.

$$\Lambda_{\mathrm{R}}^{\star} = \frac{(|S_{tb}^{(1)}|^{v_{1/2}}|S_{tb}^{(2)}|^{v_2/2}}{\left|\frac{v_1 S_{tb}^{(1)} + v_2 S_{tb}^{(2)}}{v_1+v_2}\right|^{(v_1+v_2)/2}}, \tag{4.3}$$

where $v_i$'s are the empirical degrees of freedom. When the null hypothesis $\mathrm{H}_0 : \Sigma_1 = \Sigma_2$ is true the distribution of $-2\rho \log \Lambda_{\mathrm{R}}^{\star}$ can be expanded for large $M = \rho n$ by Eq. (4.2) Of course the values of $\gamma$ and $M$ depend on $v_i$ instead of $n_i$, where $i = 1, 2$.

### 4.2. Results

First, we obtained the upper 5% and 1% critical values for different combinations of $p \in (2, 4, 6)$ and $(n_1, n_2) \in \{(20, 20), (20, 50), (100, 50), (500, 500)\}$ using the approximate equation (4.2). It was found that the critical values are very comparable for the three methods. In higher dimension the critical values get close for these three estimates as the sample size gets large.

Table 1
The table contains critical values using the simulated degrees of freedom for ($S_{tb}$), the empirical degrees of freedom ($\tilde{S}_{tb}$)

| $(n_1, n_2)$ | $\alpha$ | Estimates | $p$ | | |
|---|---|---|---|---|---|
| | | | 2 | 4 | 6 |
| (20,20) | 1% | $S_{tb}$ | 13.572 | 28.488 | 49.148 |
| | | $\tilde{S}_{tb}$ | 11.895 | 26.400 | 45.890 |
| | | $S$ | 12.023 | 26.243 | 47.118 |
| | 5% | $S_{tb}$ | 9.476 | 22.473 | 41.222 |
| | | $\tilde{S}_{tb}$ | 8.560 | 20.947 | 38.584 |
| | | $S$ | 8.286 | 20.698 | 39.522 |
| (20,50) | 1% | $S_{tb}$ | 13.084 | 26.968 | 46.322 |
| | | $\tilde{S}_{tb}$ | 11.783 | 24.399 | 46.322 |
| | | $S$ | 11.845 | 25.407 | 44.760 |
| | 5% | $S_{tb}$ | 9.017 | 21.252 | 38.842 |
| | | $\tilde{S}_{tb}$ | 8.137 | 19.250 | 34.790 |
| | | $S$ | 8.160 | 20.034 | 37.540 |
| (100,50) | 1% | $S_{tb}$ | 11.895 | 24.520 | 41.873 |
| | | $\tilde{S}_{tb}$ | 11.572 | 23.782 | 40.150 |
| | | $S$ | 11.541 | 24.034 | 41.029 |
| | 5% | $S_{tb}$ | 8.195 | 19.340 | 35.134 |
| | | $\tilde{S}_{tb}$ | 7.975 | 18.761 | 33.693 |
| | | $S$ | 7.951 | 18.958 | 34.430 |
| (500,500) | 1% | $S_{tb}$ | 11.405 | 23.362 | 39.268 |
| | | $\tilde{S}_{tb}$ | 11.384 | 23.313 | 39.154 |
| | | $S$ | 11.367 | 23.307 | 39.179 |
| | 5% | $S_{tb}$ | 7.857 | 18.429 | 32.954 |
| | | $\tilde{S}_{tb}$ | 7.843 | 18.390 | 32.859 |
| | | $S$ | 7.831 | 18.386 | 32.880 |

Classical LR statistic is used for computing values corresponding to sample covariance ($S$).

To examine the performance of the robust LR statistic for uncontaminated data we drew 1000 Monte Carlo data sets from $N_p(0, I)$ for each of the different combinations of $(n_1, n_2) \in \{(20, 20), (20, 50), (100, 50), (500, 500)\}$ and $p \in (2, 4, 6)$. The S-estimates of the covariance matrix and corresponding degrees of freedom were obtained for each sample. Table 2 contains the simulated type-I error probabilities using the respective critical values from Table 1.

The coverage probabilities corresponding to degrees of freedom $n_1 - 1$ and $n_2 - 1$ and corresponding to degrees of freedom $v_1$ and $v_2$ are very comparable. No specific pattern was observed with large *n* and *p*. This means the robust LR statistic based on S-estimates performs equally well as the non-robust one. No significant difference was observed in the performance of the robust LR statistic when empirical degrees of freedom were used. This implies we do not need to always simulate the degrees of freedom to approximate the

Table 2
Each entry corresponding to $S_{tb}$ and $\tilde{S}_{tb}$ is computed using the robust version of LR statistic with simulated degrees of freedom and with empirical degrees of freedom, respectively. Each entry corresponding to $S$ is computed by the classical LR statistic

| $(n_1, n_2)$ | $\alpha$ | Estimates | $p$ | | |
|---|---|---|---|---|---|
| | | | 2 | 4 | 6 |
| (20,20) | 1% | $S_{tb}$ | 0.024 | 0.058 | 0.017 |
| | | $\tilde{S}_{tb}$ | 0.050 | 0.058 | 0.012 |
| | | $S$ | 0.019 | 0.009 | 0.009 |
| | 5% | $S_{tb}$ | 0.074 | 0.138 | 0.057 |
| | | $\tilde{S}_{tb}$ | 0.110 | 0.139 | 0.046 |
| | | $S$ | 0.059 | 0.043 | 0.046 |
| (20,50) | 1% | $S_{tb}$ | 0.025 | 0.048 | 0.024 |
| | | $\tilde{S}_{tb}$ | 0.039 | 0.048 | 0.016 |
| | | $S$ | 0.007 | 0.006 | 0.010 |
| | 5% | $S_{tb}$ | 0.071 | 0.116 | 0.177 |
| | | $\tilde{S}_{tb}$ | 0.103 | 0.110 | 0.063 |
| | | $S$ | 0.046 | 0.042 | 0.041 |
| (100,50) | 1% | $S_{tb}$ | 0.013 | 0.015 | 0.019 |
| | | $\tilde{S}_{tb}$ | 0.012 | 0.010 | 0.019 |
| | | $S$ | 0.008 | 0.012 | 0.013 |
| | 5% | $S_{tb}$ | 0.064 | 0.055 | 0.062 |
| | | $\tilde{S}_{tb}$ | 0.060 | 0.042 | 0.066 |
| | | $S$ | 0.052 | 0.050 | 0.067 |
| (500,500) | 1% | $S_{tb}$ | 0.012 | 0.008 | 0.010 |
| | | $\tilde{S}_{tb}$ | 0.011 | 0.002 | 0.008 |
| | | $S$ | 0.012 | 0.008 | 0.010 |
| | 5% | $S_{tb}$ | 0.043 | 0.033 | 0.038 |
| | | $\tilde{S}_{tb}$ | 0.041 | 0.029 | 0.038 |
| | | $S$ | 0.061 | 0.047 | 0.048 |

Critical values are used from Table 1.

distribution of the S-estimates of $\Sigma$ for any real data set. The empirical degrees of freedom perform equally well.

The reason we developed a robust test statistic is to deal with the situation when the underlying assumption of normality for using the Bartlett test statistic is not met. Therefore, we would like to examine next the performance of the RLRT when the underlying data set is not exactly normal. Again 1000 Monte Carlo data sets were drawn from contaminated normal for different combinations of $(n_1, n_2) \in \{(20, 20), (20, 50), (100, 50), (500, 500)\}$ and $p \in (2, 4, 6)$. Each data set of size $n$ consists of 90% data from $N_p(0, I)$ and 10% data from $N_p(0, 5I)$. Table 3 contains the simulated rejection probabilities for RLRT using simulated and empirical degrees of freedom and the simulated rejection probabilities for the classical LRT. It is observed that the LRT performance becomes worse with large $p$, whereas

Table 3
This table contains simulated rejection probabilities at nominal $\alpha = 1\%$ and 5% using the respective critical values in Table 1 when the underlying population is contaminated normal, i.e. 90% of a sample is from $N_p(0, I)$ and 10% from $N_p(0, 5I)$

| $n$ | $\alpha$ | Estimates | $p$ | | |
|---|---|---|---|---|---|
| | | | 2 | 4 | 6 |
| (20,20) | 1% | $S_{tb}$ | 0.013 | 0.074 | 0.032 |
| | | $\tilde{S}_{tb}$ | 0.022 | 0.074 | 0.020 |
| | | $S$ | 0.461 | 0.822 | 0.934 |
| | 5% | $S_{tb}$ | 0.056 | 0.162 | 0.112 |
| | | $\tilde{S}_{tb}$ | 0.093 | 0.162 | 0.065 |
| | | $S$ | 0.627 | 0.915 | 0.975 |
| (20,50) | 1% | $S_{tb}$ | 0.015 | 0.053 | 0.037 |
| | | $\tilde{S}_{tb}$ | 0.022 | 0.049 | 0.024 |
| | | $S$ | 0.499 | 0.880 | 0.983 |
| | 5% | $S_{tb}$ | 0.053 | 0.131 | 0.116 |
| | | $\tilde{S}_{tb}$ | 0.074 | 0.125 | 0.084 |
| | | $S$ | 0.682 | 0.954 | 0.997 |
| (100,50) | 1% | $S_{tb}$ | 0.007 | 0.022 | 0.064 |
| | | $\tilde{S}_{tb}$ | 0.007 | 0.012 | 0.026 |
| | | $S$ | 0.544 | 0.945 | 0.998 |
| | 5% | $S_{tb}$ | 0.063 | 0.070 | 0.085 |
| | | $\tilde{S}_{tb}$ | 0.059 | 0.050 | 0.081 |
| | | $S$ | 0.693 | 0.975 | 0.999 |
| (500,500) | 1% | $S_{tb}$ | 0.006 | 0.012 | 0.013 |
| | | $\tilde{S}_{tb}$ | 0.006 | 0.012 | 0.013 |
| | | $S$ | 0.559 | 0.943 | 0.999 |
| | 5% | $S_{tb}$ | 0.037 | 0.045 | 0.050 |
| | | $\tilde{S}_{tb}$ | 0.035 | 0.043 | 0.050 |
| | | $S$ | 0.702 | 0.974 | 1.00 |

Each entry corresponding to $S_{tb}$ is based on simulated degrees of freedom. Each entry corresponding to $\tilde{S}_{tb}$ is based on empirical degrees of freedom.

the RLRT behaves very well compared to the LRT, and its performance does not vary with large values of $p$ and $n$. The results from the empirical degrees of freedom compare well with the simulated results. This paper has presented the robustness of RLRT only under 5% contamination to avoid the paper to become too lengthy. The author found quite similar results on other percentages of contamination.

## 5. Conclusion

Testing the dispersion matrix in multivariate analysis is not an easy task in classical statistics even if the distributional assumptions are met. It becomes even harder if the distributional assumptions are not satisfied and these are suspected outliers in the data. Because of high robust qualities, the S-estimators were considered as a suitable replacement for the sample variance. We assumed that the S-estimators of the covariance matrix under the multivariate normal distribution are approximately distributed as Wishart with degrees of freedom that depend on the dimension of the data and the sample size. A nonlinear formula was developed to approximate the degrees of freedom for the proposed Wishart distribution. This formula is an adjustment of the asymptotic degrees of freedom using the sample size and the dimension of a data set for an approximate degree of freedom. The two sample RLRT can easily be generalized for the *r* sample case. These two RLRTs give values similar to the respective LRTs if the data are uncontaminated. The type-I error of both the RLRT is highly insensitive to the distributional assumption irrespective of size of the sample and dimension.

## Appendix A. Simulation results

The critical values using $N_p(0, I)$ for testing the hypothesis $\mathrm{H}_0 : \Sigma_1 = \Sigma_2$ are given in Table 1.

Simulated rejection probabilities for testing the hypothesis $\mathrm{H}_0 : \Sigma_1 = \Sigma_2$ are given in Table 2.

Simulated rejection probabilities under the null hypothesis $\mathrm{H}_0 : \Sigma_1 = \Sigma_2$ when the underlying populations are contaminated are given in Table 3.

## References

Aslam, S., Rocke, D., 2003. Robust testing procedures for univariate scale parameters. Unpublished article.

Bartlett, M.S., 1937. Properties of sufficiency and statistical tests. Proc. Roy. Soc. London A 160, 268–282.

Davies, P., 1987. Asymptotic behavior of S-estimators of multivariate location parameters and dispersion matrices. Ann. Statist. 15, 1269–1292.

Grübel, R., Rocke, D., 1990. On the cumulants of affine equivariant estimators in elliptical families. J. Multivariate Anal. 35, 203–222.

Hardin, J., 2000. Multivariate outlier detection and robust clustering with minimum covariance determinant estimation and S-estimation. Ph.D. Thesis, UC Davis. Ann. Statist. 19, 2102–2119.

Lopuhaä, H., 1989. On the relation between S-estimators and M-estimators of multivariate location and covariance. Ann. Statist. 17, 1662–1683.

Mardia, K., Kent, J., Bibby, J., 1997. Multivariate Analysis. Academic Press, New York.

Muirhead, R.J., 1982. Aspects of Multivariate Statistical Theory. Wiley, New York.

Perlman, M.D., 1980. Unbiasedness of the likelihood ratio tests for equality of several covariance matrices and equality of several multivariate normal populations. Ann. Statist. 8, 247–263.

Pitman, E.J.G., 1939. Tests of hypotheses concerning location and scale parameters. Biometrika 31, 200–215.

Rocke, D., 1996. Robustness properties of S-estimators of multivariate location and shape in high dimension. Ann. Statist. 24, 1327–1345.

Rousseeuw, P., Yohai, V., 1984. Robust regression by means of S-estimators. Robust and Nonlinear Time Series Analysis. Lecture Notes in Statistics, Vol. 26, Springer, New York, pp. 256–272.

Tyler, D., 1983. Robust and efficiency properties of scatter matrices. Biometrika 70, 411–420.