Statistica Neerlandica (1993) Vol. 47, nr. 3, pp. 199-219

Testing equality of variances in the analysis of repeated measurements

W. C. M. Kallenberg
Faculty of Applied Mathematics, University of Twente
P.O. Box 217, 7500 AE Enschede, The Netherlands

G. S. Mudholkar
Statistics Department, University of Rochester
Rochester, NY 14627, U.S.A.

P. Subbaiah
Department of Mathematical Sciences, Oakland University
Rochester, MI 48309, U.S.A.

The problem of comparing the precisions of two instruments using repeated measurements can be cast as an extension of the Pitman-Morgan problem of testing equality of variances of a bivariate normal distribution. Hawkins (1981) decomposes the hypothesis of equal variances in this model into two subhypotheses for which simple tests exist. For the overall hypothesis he proposes to combine the tests of the subhypotheses using Fisher's method and empirically compares the component tests and their combination with the likelihood ratio test. In this paper an attempt is made to resolve some discrepancies and puzzling conclusions in Hawkins's study and to propose simple modifications.

The new tests are compared to the tests discussed by Hawkins and to each other both in terms of the finite sample power (estimated by Monte Carlo simulation) and theoretically in terms of asymptotic relative efficiencies.

Keywords & Phrases: repeated measurements, equality of variances, combination of tests, Monte Carlo simulation, asymptotic efficiency.

1 Introduction

The problem of testing equality of the variances of two correlated normal variables was resolved independently and almost simultaneously by Pitman (1939) and Morgan (1939). The likelihood ratio procedure obtained by them consists of testing independence of two normal variables. Hawkins (1981) considered this problem in the context of the model

$$X_{ij} = \xi_i + \delta_j + \varepsilon_{ij}, \tag{1}$$

$i = 1, 2, ..., n; j = 1, 2, ..., p + q$, where $(X_{i1}, ..., X_{ip}, X_{i\,p+1}, ..., X_{i\,p+q})$ are the observations on the i^{th} subject obtained using two instruments; the first p components with instrument 1, and the next q components with instrument 2. The measurements X_{ij} may alternatively be the responses from the i^{th} subject after receiving two drugs, the first p responses being associated with drug 1 and the next q with drug 2. It is assumed that the ε_{ij}'s are independent and normally distributed with zero means and variances given by

$$\operatorname{Var}(\varepsilon_{ij}) = \begin{cases} \sigma_1^2, j = 1, 2, \ldots, p \\ \sigma_2^2, j = p+1, \ldots, p+q. \end{cases}$$

Note that in the preceding example on a clinical drug trial a possible serial correlation between the readings taken on the same subject at different times is neglected.

The parameters δ_j are the fixed effects associated with the treatments, $\delta_j = \delta_{(1)}$, $j = 1, \ldots, p$ and $\delta_j = \delta_{(2)}, j = p+1, \ldots, p+q$; the random variables $\xi_1, \ldots, \xi_n$ are independent and normally distributed with the same expectation, which w.l.o.g. may be taken equal to 0, and variance τ^2. It is assumed that

$$\sigma_1^2 \geq 0, \quad \sigma_2^2 \geq 0, \quad \tau^2 \geq 0, \quad \sigma_1^2 + \sigma_2^2 > 0, \quad \sigma_1^2 + \tau^2 > 0, \quad \sigma_2^2 + \tau^2 > 0, \tag{2}$$

where a normal distribution with variance 0 corresponds to a degenerate distribution. In particular, the random variables ξ_i may be degenerate.

Further, the ξ_i's and ε_{ij}'s are independent. Since the primary concern in the evaluation of the instruments is their precision, a major statistical problem in this model is that of testing $H_0: \sigma_1^2 = \sigma_2^2$. It is easy to see that when $p = q = 1$, the problem reduces to a problem closely related to that considered by PITMAN and MORGAN. (Note that there are some restrictions on the covariance matrix, due to $\tau^2 \geq 0$).

The assumptions of model (1) imply that the covariance matrix of X_i, the vector of observations for the i^{th} individual, is

$$\operatorname{Var}(X_i) = \begin{bmatrix} \sigma_1^2 I_p + \tau^2 E_{pp} & \tau^2 E_{pq} \\ \tau^2 E_{qp} & \sigma_2^2 I_q + \tau^2 E_{qq} \end{bmatrix},$$

where I_m denotes the identity matrix of order m, and E_{mn} denotes the $m \times n$ matrix of unities. The structure of the problem is simplified by making orthogonal transformations of the p observations with treatment 1 and the q observations with treatment 2 as follows.

Let H_m denote the Helmert matrix with $(m^{-1/2}, \ldots, m^{-1/2})$ as the first row, and let

$$Y_i = \begin{bmatrix} H_p & 0 \\ 0 & H_q \end{bmatrix} X_i, i = 1, \ldots, n.$$

Then it is easily verified that

$$\operatorname{Var}(Y_i) = \begin{bmatrix} \sigma_1^2 + p\tau^2 & 0 & \tau^2 (pq)^{1/2} & 0 \\ 0 & \sigma_1^2 I_{p-1} & 0 & 0 \\ \tau^2 (pq)^{1/2} & 0 & \sigma_2^2 + q\tau^2 & 0 \\ 0 & 0 & 0 & \sigma_2^2 I_{q-1} \end{bmatrix}.$$

Furthermore the components of Y_i may be rearranged so that

$$\operatorname{Var}\begin{bmatrix} Y_{i1} \\ Y_{i\,p+1} \\ \cdots \\ Y_{i2} \\ \vdots \\ Y_{i\,p} \\ \cdots \\ Y_{i\,p+2} \\ \vdots \\ Y_{i\,p+q} \end{bmatrix} = \begin{bmatrix} \sigma_1^2 + p\tau^2 & \tau^2(pq)^{1/2} & 0 & 0 \\ \tau^2(pq)^{1/2} & \sigma_2^2 + q\tau^2 & 0 & 0 \\ 0 & 0 & \sigma_1^2 I_{p-1} & 0 \\ 0 & 0 & 0 & \sigma_2^2 I_{q-1} \end{bmatrix}$$

with corresponding expected value $(\sqrt{p}\delta_{(1)}, \sqrt{q}\delta_{(2)}, 0, \ldots, 0)$.

The problem is discussed in detail by Hawkins for the case $p = q$. A second transformation

$$\begin{aligned} &Z_{i1} = (Y_{i1} + Y_{i\,p+1})/\sqrt{2}, \quad Z_{i2} = (Y_{i1} - Y_{i\,p+1})/\sqrt{2}, \\ &Z_{i3} = Y_{i2}, \ldots, Z_{i\,p+1} = Y_{i\,p}, \quad Z_{i\,p+2} = Y_{i\,p+2}, \ldots, Z_{i2p} = Y_{i2p} \end{aligned} \tag{3}$$

yields normal variables with covariance matrix

$$\operatorname{Var}(Z_i) = \begin{bmatrix} \frac{1}{2}(\sigma_1^2 + \sigma_2^2) + 2p\tau^2 & \frac{1}{2}(\sigma_1^2 - \sigma_2^2) & 0 & 0 \\ \frac{1}{2}(\sigma_1^2 - \sigma_2^2) & \frac{1}{2}(\sigma_1^2 + \sigma_2^2) & 0 & 0 \\ 0 & 0 & \sigma_1^2 I_{p-1} & 0 \\ 0 & 0 & 0 & \sigma_2^2 I_{p-1} \end{bmatrix}. \tag{4}$$

For the general case Hawkins does not identify a transformation that simplifies the problem. However, we note that the transformation

$$\left.\begin{aligned} &Z_{i1} = (\sqrt{p}Y_{i1} + \sqrt{q}Y_{i\,p+1})/\sqrt{p+q}, \quad Z_{i2} = (\sqrt{q}Y_{i1} - \sqrt{p}Y_{i\,p+1})/\sqrt{p+q} \\ &Z_{i3} = Y_{i2}, \ldots, Z_{i\,p+1} = Y_{i\,p}, \quad Z_{i\,p+2} = Y_{i\,p+2}, \ldots, Z_{i\,p+q} = Y_{i\,p+q} \end{aligned}\right\} \tag{5}$$

leads to normal variables with covariance matrix

$$\operatorname{Var}(Z_i) = \begin{bmatrix} \dfrac{p\sigma_1^2 + q\sigma_2^2}{p+q} + (p+q)\tau^2 & \dfrac{(\sigma_1^2 - \sigma_2^2)(pq)^{1/2}}{p+q} & 0 & 0 \\ \dfrac{(\sigma_1^2 - \sigma_2^2)(pq)^{1/2}}{p+q} & \dfrac{(q\sigma_1^2 + p\sigma_2^2)}{p+q} & 0 & 0 \\ 0 & 0 & \sigma_1^2 I_{p-1} & 0 \\ 0 & 0 & 0 & \sigma_2^2 I_{q-1} \end{bmatrix} \tag{6}$$

When $p = q$, (5) and (6) reduce to (3) and (4), respectively.

2 Tests by Hawkins

For testing $H_0: \sigma_1^2 = \sigma_2^2$, when $p = q$, Hawkins (1981) examines the likelihood ratio test labelled GLR (generalized likelihood ratio) in terms of the Z_{ij}'s. The GLR while feasible gives serious problems. It requires numerical maximization of the likelihood function. Moreover, its null distribution which involves the nuisance parameter τ^2 is unavailable. For examining the power function of the GLR, simulated percentiles are obtained for various values of the nuisance parameter.

We can not even use standard asymptotical theory. As is shown in Appendix A, the maximum likelihood estimator of τ^2 is irregular under $\tau^2 = 0$ ($\tau^2 = 0$ is advised by HAWKINS to take in practice and is called "worst-case"). Note also the correction on p. 46 of HAWKINS (1981) to the asymptotic χ_1^2-theory if $\tau^2 = 0 : \hat{\alpha}_0 = 0.08$, while $\hat{\alpha}_1 = \hat{\alpha}_4 = 0$. So, the GLR is difficult to apply and it is worthwile to develop other test statistics, which

(i) should be easy to compute,

(ii) should have an easy (asymptotic) null distribution, and

(iii) should be nearly as powerful as the GLR.

In an attempt to develop an alternative to the GLR, it is noted that the testing problem can be decomposed in two more simple subproblems, testing (i) $H_{01}: \text{cov}(Z_{i1}, Z_{i2}) = 0$, and (ii) $H_{02}: (Z_{i3}, ..., Z_{i\,p+1})$ and $(Z_{i\,p+2}, ..., Z_{i2p})$ have equal variances. The likelihood ratio test of H_{01} due to Pitman-Morgan, based on (Z_{i1}, Z_{i2}), $i = 1, ..., n$, ignoring the restrictions induced by $\tau^2 \geq 0$, reduces to a t-test for zero correlation between Z_{i1} and Z_{i2}. For the hypothesis H_{02} of equal variances the alternative is two sided for which HAWKINS uses an F-test. The corresponding test statistics T and F are independent and the "pooled test" for H_0 obtained by combining the two P-values using Fisher's method provides an alternative to the GLR.

The power functions of T, F, the pooled test and the GLR are compared using a Monte Carlo experiment. From the empirical study it is concluded that (i) the GLR is the best of the four tests and its power decreases as τ^2 increases; (ii) T alone is the poorest, (iii) the power of the GLR decreases to that of F as τ^2 increases; (iv) the combination of T and F is weaker occasionally than one of its components namely the F-test. The pooled test seems better when $\tau^2 < 1$, and the F-test seems better when $\tau^2 \geq 1$. HAWKINS then suggests estimating τ^2 and using it to decide between F and the pooled test. The value 1 here is due to the choice $\sigma_1^2 = \sigma_2^2 = 1$ (under H_0) in the Monte Carlo experiment. Not the magnitude of τ^2 itself, but of τ^2/σ_i^2 is important.

In this paper we try to explain the role of τ^2/σ_i^2 in the preceding conclusions. Moreover, we present a new test statistic, which is asymptotically optimal for testing the conjunction of H_{01} and H_{02}, which is easy to compute and has an easy asymptotic null-distribution: the standard-normal distribution. It turns out that the new test is as powerful as GLR, even for $n = 10$. So this test satisfies the three aims mentioned before. *Therefore we recommend this statistic, given in (7) in section 3, for testing* $\sigma_1^2 = \sigma_2^2$ *against* $\sigma_1^2 \neq \sigma_2^2$. This new test statistic may be seen as a weighted combination of T and F.

Further we consider several classical combination procedures for testing H_{01} and H_{02}. We also consider another way of decomposing the testing problem, H_{01} and H_{02}^*, say. We discuss tests for H_{02}^* only, and several combination procedures.

In section 3 the various modified tests are described. In section 4 we estimate the power functions of these modified procedures by simulation with the parametric configuration considered by HAWKINS. In section 5 the tests are theoretically compared in terms of Pitman efficiencies and (local) slopes. Also some open questions and some suggestions are presented.

3 Modified tests

In this section we examine Hawkins' study and propose some new tests. Among the conclusions of Hawkins' simulation study the optimality of the GLR, the general strength of the F-test and the weakness of T alone are reasonable. At first sight it looks rather strange that the combination of T and F (where one uses more information) is occasionally weaker than the F-test alone. To clarify this point, the role of τ^2/σ_i^2 has to be investigated more closely.

If τ^2 is large with respect to σ_1^2 and σ_2^2, the power of the T-test at the point $(\sigma_1^2, \sigma_2^2, \tau^2)$ will be low, since the power is low if the correlation between Z_{i1} and Z_{i2} is low and this is the case when τ^2 is large. In other words the T-test is not very informative for large values of τ^2 (with respect to σ_1^2 and σ_2^2). This explains why the F-test alone is sometimes (for τ^2 large!) more powerful than the pooled test, because in that case applying Fisher's method the F-test is combined in an unweighted way with a poor test. Nevertheless, it should be possible to use both the information in the T-test and in the F-test. Although in general Fisher's method of combining independent tests may be effective, in this situation it seems to be better to combine the T-test and F-test in a different way, using their asymptotic normal distributions. In view of the bad performance of the T-test for large values of τ^2 it seems to be more appropriate to give the T-test a *low* weight if τ^2 is relatively large (cf. also Hawkins' proposal to estimate τ^2 and then to select the pooled or F-test; here we consider a more refined method of weighting).

We now express the test statistics in terms of the $Z_i = (Z_{i1}, Z_{i2}, \ldots, Z_{i\,p+q})'$, $i = 1, 2, \ldots, n$ and use the form of the dispersion matrix of Z_i, given by (6). Since $E(Z_{i1})$, $E(Z_{i2})$ involve unknown parameters and $E(Z_{ij}) = 0$, $j = 3, \ldots, p+q$, we compute the sample covariance matrix of $Z_1, \ldots, Z_n$ as $S^Z = ((S_{jk}^Z))$ where

$$S_{jk}^Z = \begin{cases} \sum_{i=1}^{n} (Z_{ij} - \bar{Z}_{.j})(Z_{ik} - \bar{Z}_{.k})/(n-1), & \text{for } j,k = 1,2 \\ \sum_{i=1}^{n} Z_{ij} Z_{ik}/n, & \text{otherwise} \end{cases}$$

and

$$\bar{Z}_{.j} = \sum_{i=1}^{n} Z_{ij}/n, \quad j = 1, \ldots, p+q.$$

The null hypothesis $H_0: \sigma_1^2 = \sigma_2^2$ is now replaced by two subhypotheses, $H_{01}: Z_{i1}$ and Z_{i2} are independent, and $H_{02}: Z_{i3}, \ldots, Z_{i\,p+1}$, and $Z_{i\,p+2}, \ldots, Z_{i\,p+q}$ have the same variances, which may be tested using

$$T = R_{12}[(n-2)/(1-R_{12}^2)]^{1/2},$$
$$F = S_3^2/S_4^2,$$

where

$$R_{12} = S_{12}^Z/(S_{11}^Z S_{22}^Z)^{1/2}, \quad S_3^2 = \sum_{i=3}^{p+1} S_{ii}^Z/(p-1), \quad S_4^2 = \sum_{i=p+2}^{p+q} S_{ii}^Z/(q-1).$$

Note that T^2 with T as defined above corresponds to Hawkins' T. Further note that the test statistic T is easy to compute also in case $p \neq q$ in spite of Hawkins' (1981) remark on p. 45.

Under H_0, T follows Student's t-distribution with $(n-2)$ df, and F follows Snedecor's F-distribution with $(n(p-1), n(q-1))$ df.

Next consider contiguous alternatives of the form

$$\frac{\sigma_1^2}{\sigma_2^2} = 1 + \theta n^{-1/2}.$$

To derive the asymptotically optimal weights, first suppose τ^2/σ_2^2 to be known. Defining

$$W = \frac{\sqrt{n-3}}{2} \log\left(\frac{1+R_{12}}{1-R_{12}}\right)$$

it follows that

$$W \xrightarrow{D} N(\theta c_1, 1) \quad \text{as} \quad n \to \infty$$

with

$$c_1 = (pq)^{1/2}(p+q)^{-1}(1+\tau^2(p+q)(\sigma_2^2)^{-1})^{-1/2}.$$

(For the norming factor $\sqrt{n-3}$ see, for example, JOHNSON and KOTZ, 1970b, p. 229.) Note that W and T are equivalent test statistics for testing H_{01}. Here again it is seen that we obtain low power if τ^2 is relatively large, since in that case c_1 is relatively small.

Based on the Wilson-Hilferty approximation to the distribution of chi-squared random variables, cf. JOHNSON and KOTZ, (1970b), p. 83, define

$$V = \left[\left\{1 - \frac{2}{9n(q-1)}\right\} F^{1/3} - \left\{1 - \frac{2}{9n(p-1)}\right\}\right] \left[\frac{2}{9n(q-1)} F^{2/3} + \frac{2}{9n(p-1)}\right]^{-1/2}.$$

It follows that

$$V \xrightarrow{D} N(\theta c_2, 1) \quad \text{as} \quad n \to \infty$$

with

$$c_2 = \frac{(p-1)^{1/2}(q-1)^{1/2}}{2^{1/2}(p+q-2)^{1/2}}.$$

So at first order the two testing problems together are equivalent to testing $\theta = 0$ against $\theta \neq 0$ with two independent normal variables with variance 1 and expectation θc_1 and θc_2, respectively. Now it is clear how we have to pool both tests, rejecting H_0 if

$$\left|\frac{c_1 W + c_2 V}{\sqrt{c_1^2 + c_2^2}}\right| > \Phi^{-1}(1 - \tfrac{1}{2}\alpha),$$

where Φ denotes the standard-normal distribution function and α the (approximate) level of the test. Note that if τ^2 is relatively large the weight c_1 of W is small, which agrees with the above reasoning. The final step is replacing c_1 by $\hat{c}_1$, where $\hat{c}_1$ is obtained by inserting a consistent estimator of τ^2/σ_2^2 in c_1 for which we take $\hat{\tau}^2/\hat{\sigma}_2^2$ with

$$\hat{\tau}^2 = \max\{0, [S_{11}^Z - S_{22}^Z - (p-q)(pq)^{-1/2} S_{12}^Z](p+q)^{-1}\}$$

and

$$\hat{\sigma}_2^2 = S_4^2.$$

So the test statistic is

$$U = (\hat{c}_1 W + c_2 V)(\hat{c}_1^2 + c_2^2)^{-1/2} \tag{7}$$

and we reject H_0 at level α if $|U| > \Phi^{-1}(1 - \frac{1}{2}\alpha)$.

REMARK 1. Although simulation results show that even for $n = 10$ the actual level of the test based on U is close to the nominal level α, a modification is possible using test statistics, which are exactly $N(0, 1)$-distributed under H_0. The idea of this modification is due to the referee. It can be worked out as follows. The one-sided P-value corresponding with T is defined by

$$Q_1(T) \quad \text{with} \quad Q_1(t) = Pr(T_{n-2} \leq t),$$

where T_{n-2} has a t-distribution with $(n-2)$ df.

Under H_0, $Q_1(T)$ is uniformly distributed on $(0, 1)$ and therefore $\Phi^{-1}[Q_1(T)]$ has a $N(0, 1)$-distribution under H_0. Under contiguous alternatives of the form $\sigma_1^2 \sigma_2^{-2} = 1 + \theta n^{-1/2}$ we get

$$\Phi^{-1}[Q_1(T)] \xrightarrow{D} N(\theta c_1, 1) \quad \text{as} \quad n \to \infty.$$

The one-sided P-value corresponding with F is defined by

$$Q_2(F) \quad \text{with} \quad Q_2(f) = Pr\{F(n(p-1), n(q-1)) \leq f\},$$

where $F[n(p-1), n(q-1)]$ has a F-distribution with $[n(p-1), n(q-1)]$df. Under H_0, $\Phi^{-1}[Q_2(F)]$ has a $N(0, 1)$-distribution, while under contiguous alternatives of the form $\sigma_1^2 \sigma_2^{-2} = 1 + \theta n^{-1/2}$ we have

$$\Phi^{-1}[Q_2(F)] \xrightarrow{D} N(\theta c_2, 1) \quad \text{as} \quad n \to \infty.$$

Arguing as before we may replace in (7) W by the equivalent test statistic $W^* = \Phi^{-1}[Q_1(T)]$ and V by the equivalent test statistic $V^* = \Phi^{-1}[Q_2(F)]$, rejecting H_0 if

$$|(\hat{c}_1 W^* + c_2 V^*)(\hat{c}_1^2 + c_2^2)^{-1/2}| > \Phi^{-1}(1 - \tfrac{1}{2}\alpha).$$

Now W^* and V^* are *exactly* $N(0, 1)$-distributed under H_0. Note that we do not combine the two-sided P-values, but first combine the one-sided P-values and then take the two-sided critical region (as we also combine W and V in (7) and not $|W|$ and $|V|$). This is due to the fact that the two-sided P-values have a less nice limiting distribution under (contiguous) alternatives.

Of course, calculating W^* and V^* might be slightly less easier than calculating W and V.

The tests based on T, F, U and Fisher's combination of T and F all fall into the following scheme.

- The testing problem is decomposed in two more simple subproblems.
- For each subproblem one or more suitable tests are chosen.
- The two tests are combined or a test for only one of the two subproblems is considered.

Next a lot of other tests will be discussed along the preceding scheme. An examination of the structure of the covariance matrix (6) suggests an alternative way of decomposing $H_0: \sigma_1^2 = \sigma_2^2$, namely in $H_{01}: Z_{i1}$ and Z_{i2} are independent, and $H_{02}^*: Z_{i2}, Z_{i3}, \ldots, Z_{i\,p+1}, Z_{i\,p+2}, \ldots, Z_{i\,p+q}$ have the same variances, cf. REMARK 3 (section 5). So, three subproblems are involved and now we will discuss in a systematic way suitable tests for each of these three subproblems.

Testing H_{01} against the alternative $\sigma_1^2 \neq \sigma_2^2$ is close to testing independence of Z_{i1} and Z_{i2} against dependence (the difference is caused by the restriction on τ^2). The test statistic T is a standard test statistic for the latter testing problem and therefore T seems suitable for testing H_{01}.

For testing H_{02} both F and Bartlett's statistic B_2 are standard test statistics. Bartlett's statistic B_2 is defined by

$$B_2 = M_2/b_2 \quad \text{with}$$

$$M_2 = n(p+q-2)[\log\{aS_3^2 + (1-a)S_4^2\} - \{a \log S_3^2 + (1-a)\log S_4^2\}],$$

$$a = \frac{p-1}{p+q-2}, \quad b_2 = 1 + \frac{1}{3n(p+q-2)}\left\{\frac{1}{a} + \frac{1}{1-a} - 1\right\}.$$

For testing H_{02}^* we apply Bartlett's statistic B_3, defined by

$$B_3 = M_3/b_3 \quad \text{with}$$

$$M_3 = \{(n-1) + n(p+q-2)\}[\log\{a_1 S_{22}^Z + a_2 S_3^2 + a_3 S_4^2\} - \{a_1 \log S_{22}^Z + a_2 \log S_3^2 + a_3 \log S_4^2\}],$$

$$a_1 = \frac{n-1}{A}, \quad a_2 = \frac{n(p-1)}{A}, \quad a_3 = \frac{n(q-1)}{A}, \quad A = (n-1) + n(p+q-2)$$

$$b_3 = 1 + \frac{1}{6A}\left\{\frac{1}{a_1} + \frac{1}{a_2} + \frac{1}{a_3} - 1\right\}.$$

The test statistic B_3 is suitable if $(p+q)^{-1}(q\sigma_1^2 + p\sigma_2^2)$ in (6) is replaced by σ_3^2 and we want to test the hypothesis $\sigma_1^2 = \sigma_2^2 = \sigma_3^2$, based on S_{22}^Z, S_3^2, S_4^2.

REMARK 2. Another possibility is to test $\sigma_1^2 = \sigma_2^2 = \sigma^2$ by the F-test mentioned before, to test $(q\sigma_1^2 + p\sigma_2^2)/(p+q) = \sigma^2$ separately by using

$$F^* = (p+q-2)S_{22}^Z/\{(p-1)S_3^2 + (q-1)S_4^2\},$$

and to combine these two test-statistics. Note however, that F and F^* are independent under H_0, but not under alternatives, cf. also Remarks 4 and 5 in section 5.

We consider also two other test statistics for testing H_{02}^{*}:

$$D_1 = \sqrt{n}(S_3^2 - S_4^2)/S_{22}^Z \tag{8}$$

and

$$D_2 = \sqrt{n}\{(S_3^2)^{1/3} - (S_4^2)^{1/3}\}/(S_{22}^Z)^{1/3}. \tag{9}$$

Both test statistics D_1 and D_2 are simple to compute and their distributions do not involve τ^2. The null hypothesis is rejected when these statistics are too small or too large. The critical values associated with these statistics can be obtained either by simulation or by using Cornish-Fisher approximation for percentiles (cf. JOHNSON and KOTZ, 1970a, p. 34) involving the first four moments.

For large n it follows by COX and REID (1987) that under H_0

$$P(D_2 \leq x) = P(U_1 \leq x) + O(n^{-1}) \quad \text{as} \quad n \to \infty,$$

where U_1 has a normal distribution with mean $(2/9)n^{-1/2}\{(p-1)^{-1} - (q-1)^{-1}\}$ and variance $(2/9)\{(p-1)^{-1} + (q-1)^{-1}\}$. Writing

$$d_2' = [(2/9)\{(p-1)^{-1} + (q-1)^{-1}\}]^{1/2}\Phi^{-1}(1 - \tfrac{1}{2}\alpha)$$

we therefore have

$$P(|D_2| > d_2') = \alpha + O(n^{-1}) \quad \text{as} \quad n \to \infty,$$

thus obtaining a simple and theoretically accurate approximation of the critical value. Note however, the discussion on critical values in section 4. We conclude the description of tests for the three subproblems by noting that D_1 and D_2 are strongly related as is seen from

$$D_1 = 3\left(\frac{S_3^2 S_4^2}{(S_{22}^Z)^2}\right)^{1/3} D_2 + n^{-1}D_2^3.$$

Having discussed tests for the subproblems we now come to the combination of the tests. Combination of k independent tests can be done in several ways. First of all we have the combined statistic U and its modification, given in REMARK 1. These tests have already been discussed in detail at the beginning of this section. Next we mention some classical combination statistics based on (two-sided) P-values. If $P_1, \ldots, P_k$ are the P-values associated with k independent tests, then consider

(i) $\psi_F = \sum_{i=1}^{k} -2\log P_i$, due to FISHER,

(ii) $\psi_L = \sum_{i=1}^{k} -\log\{P_i/(1-P_i)\}$, studied by MUDHOLKAR and GEORGE (1979),

(iii) $\psi_T = -\min_{1\leq i\leq k} P_i$, due to TIPPETT.

Under H_0, ψ_F is distributed as χ^2 with $2k$ df, ψ_T is distributed as minus the first order statistic from a sample of size k from a uniform distribution, and ψ_L is distributed as the sum of k i.i.d. logistic variables which can be well approximated by $\pi[k(5k+2)/\{3(5k+4)\}]^{1/2} \times T_{5k+4}$ where T_{5k+4} denotes a random variable with Student's t-distribution with $5k+4$ df. The critical values of the combination statistics are the upper α-percentiles of these null distributions.

We apply the three methods on T and B_2, which are used for testing H_{01} and H_{02}, and on T and B_3, which are used for testing H_{01} and H_{02}^*. The P-value corresponding with T is defined by

$$P_1(T) \quad \text{with} \quad P_1(t) = 2Pr(T_{n-2} > |t|).$$

For Bartlett's statistic B_2 we use its limiting distribution.

The (approximate) P-value associated with B_2 is given by

$$P_2(B_2) \quad \text{with} \quad P_2(b) = Pr(\chi_1^2 > b).$$

Further it is immediately seen that T and B_2 are independent and hence also $P_1(T)$ and $P_2(B_2)$ are independent. The three combination statistics mentioned above are denoted by ψ_{F2}, ψ_{L2} and ψ_{T2}.

Bartlett's statistic B_3, used for testing H_{02}^*, is under H_0 approximately distributed as χ^2 with 2 df. The (approximate) P-value associated with B_3 is given by

$$P_2^*(B_3) \quad \text{with} \quad P_2^*(b) = P_2^*(\chi_2^2 > b).$$

In order to show that the P-values being combined: P_1, P_2^* are independently distributed under H_0, we need the following:

LEMMA 1. *Under $H_0: \sigma_1^2 = \sigma_2^2 = \sigma^2$, say, the statistics T and B_3 are independently distributed.*

PROOF. The null distribution of T does not depend on the nuisance parameters $\delta_{(1)}$, $\delta_{(2)}$, σ^2 and τ^2. Moreover, $(\bar{Z}_{.1}, \bar{Z}_{.2}, S_{11}^Z, S_{22}^Z)$ are the complete sufficient statistics for the nuisance parameters $(\delta_{(1)}, \delta_{(2)}, \sigma^2, \tau^2)$, when restricting attention to Z_{ij}, $i = 1, \dots, n, j = 1, 2$.

Hence, using BASU's (1955) theorem, it follows that T is independent of S_{22}^Z. It is obvious that (T, S_{22}^Z) is also independent of (S_3^2, S_4^2). Consequently, B_3 which is a function of S_{22}^Z, S_3^2, S_4^2, is independent of T. □

It follows from this lemma that P_1, P_2^*, the P-values associated with T and B_3, are independent when the null hypothesis is true.

The three combination statistics of P_1 and P_2^* are denoted by ψ_{F3}, ψ_{L3} and ψ_{T3}. In summary we have the following test statistics

- T for testing H_{01} only,
- B_2 or F for testing H_{02} only,
- B_3, D_1 or D_2 for testing H_{02}^* only,
- ψ_{F2}, ψ_{L2}, ψ_{T2} and U: combination tests for H_{01} and H_{02},
- ψ_{F3}, ψ_{L3}, ψ_{T3}: combination tests for H_{01} and H_{02}^*.

4 Comparison of power functions

In order to investigate the power functions of the tests described in section 3, a Monte Carlo study is conducted. In the simulation study we considered the same values for the parameters as in Hawkins' study, viz., $n = 10, 20, 50$; $p = q = 2, 3, 4, 5$; $\tau^2 = 0, 1, 4$. The power for the GLR is taken from HAWKINS (1981). A random sample of size n is generated from a bivariate normal population with mean 0 and covariance matrix specified by the first principal minor of Var (Z_i) in (6), using IMSL routine RNMVN and S_{11}^Z, S_{12}^Z, S_{22}^Z are calculated. Similarly, $n(p-1)$ observations from $N(0, \sigma_1^2)$, $n(q-1)$ observations from $N(0, \sigma_2^2)$ are generated with the IMSL routine RNNOVA and S_3^2, S_4^2 are calculated. Then the test statistics T, F, B_3, ψ_{F2}, ψ_{L2}, ψ_{T2}, ψ_{F3}, ψ_{L3}, ψ_{T3}, D_1, D_2 and U are obtained and compared with the corresponding critical values. This process is repeated for 5000 samples, and the power of each test is estimated by the proportion of times the test leads to rejection of the null hypothesis. The standard error for the estimate of the power is at most $(20000)^{-1/2} = 0.007$.

Calculation of critical constants:

The upper $100(\alpha/2)$-percentage point of the t-distribution with $(n-2)$ df is compared with $|T|$. In the simulation $p = q$ and hence F is used instead of B_2. The maximum of F and $1/F$ is compared with the upper $100(\alpha/2)$-percentage point of the F-distribution with $(n(p-1), n(p-1))$ df. The P-value $P_2(F)$ which is needed in calculating ψ_{F2}, ψ_{L2}, ψ_{T2} is obtained as

$$P_2(F) = 2Pr(F_{(n(p-1), n(p-1))} > \max(F, 1/F)).$$

B_3 is compared with $\chi^2_{1-\alpha;2}$, the upper 100α-percentile of the χ^2-distribution with 2 df. The critical constant for ψ_{F2}, ψ_{F3} is $\chi^2_{1-\alpha;4}$. ψ_{L2}, ψ_{L3} are compared with $\pi(24/42)^{1/2} t_{1-\alpha;14}$, where as ψ_{T2}, ψ_{T3} are compared with $-[1-(1-\alpha)^{1/2}]$. $|U|$ is compared with $z_{1-\alpha/2}$, the upper $100(\alpha/2)$-percentile of the standard normal distribution.

The critical constants for D_1 and D_2 are obtained using two methods. One simple way is as follows: For large n, $(\chi_n^2/n)^{1/3}$ is approximately normally distributed with mean $1-2(9n)^{-1}$ and variance $2(9n)^{-1}$ based on the Wilson-Hilferty approximation. Let $U_1 = n^{1/2}\sigma^{-2/3}\{(S_3^2)^{1/3} - (S_4^2)^{1/3}\}$, $U_2 = (\sigma^{-2} S_{22}^Z)^{1/3}$, and d_2 be a constant such that under H_0, $Pr(D_2 = U_1/U_2 \le d_2) = 1-\alpha$, i.e., $Pr(U_1 - d_2 U_2 \le 0) = 1-\alpha$. Since U_1 is approximately normal with mean 0, and variance $\dfrac{4}{9(p-1)}$, and U_2 is approximately normal with mean $1 - \dfrac{2}{9(n-1)}$, and variance $\dfrac{2}{9(n-1)}$, d_2 is obtained as

$$d_2 = \left[\frac{(z_{1-\alpha/2})^2 \dfrac{4}{9(p-1)}}{\left[1 - \dfrac{2}{9(n-1)}\right]^2 - (z_{1-\alpha/2})^2 \dfrac{2}{9(n-1)}} \right]^{1/2}.$$

Note that as $n \to \infty$, $d_2 \to d_2' = (z_{1-\alpha/2}) \left[\frac{4}{9(p-1)}\right]^{1/2}$, the critical constant based on the approximation, given in section 3. The power using the constants d_2 and d_2' is denoted by column headings D_2 and D_2' in Table 1. Similarly the critical constants for D_1 are

$$d_1 = \left[\frac{(z_{1-\alpha/2})^2 \dfrac{4}{(p-1)}}{1-(z_{1-\alpha/2})^2 \dfrac{2}{(n-1)}}\right]^{1/2} \quad \text{and} \quad d_1' = z_{1-\alpha/2}\left[\frac{4}{(p-1)}\right]^{1/2}.$$

These critical constants are summarized in the following table.

TEST STATISTIC	CRITICAL CONSTANT
$\max(F, 1/F)$	$F_{1-\alpha/2;(n(p-1),n(p-1))}$
$\|T\|$	$t_{1-\alpha/2;(n-2)}$
B_3	$\chi^2_{1-\alpha;2}$
$\|U\|$	$z_{1-\alpha/2} = \Phi^{-1}(1-\frac{1}{2}\alpha)$
ψ_{F2}, ψ_{F3}	$\chi^2_{1-\alpha;4}$
ψ_{L2}, ψ_{L3}	$\pi(24/42)^{1/2} t_{1-\alpha;14}$
ψ_{T2}, ψ_{T3}	$-[1-(1-\alpha)^{1/2}]$
$\|D_1\|$	$d_1 = \left[\frac{(z_{1-\alpha/2})^2 \frac{4}{(p-1)}}{1-(z_{1-\alpha/2})^2 \frac{2}{(n-1)}}\right]^{1/2}$ (or) $d_1' = z_{1-\alpha/2}\left[\frac{4}{(p-1)}\right]^{1/2}$
$\|D_2\|$	$d_2 = \left[\frac{(z_{1-\alpha/2})^2 \frac{4}{9(p-1)}}{\left[1-\frac{2}{9(n-1)}\right]^2 - (z_{1-\alpha/2})^2 \frac{2}{9(n-1)}}\right]^{1/2}$ (or) $d_2' = z_{1-\alpha/2}\left[\frac{4}{9(p-1)}\right]^{1/2}$

The estimated power of the tests is given in Table 1. The actual significance levels of the tests are also estimated in the Monte Carlo study. Except for D_1, D_1' and D_2' the actual levels agree with the nominal level 0.05.

Table 1. Estimates of power function based on 5000 samples with $\alpha=.05$; $\sigma_1^2=2$, $\sigma_2^2=1$ (probabilities $\times 10^2$)

n	p	τ^2	F	B_3	D_1**	D_1'**	D_2	D_2'**	T	U	ψ_{F2}	ψ_{L2}	ψ_{T2}	ψ_{F3}	ψ_{L3}	ψ_{T3}	GLR*
10	2	0	17	14	4	25	17	20	15	29	21	22	19	19	19	17	28
10	2	1	17	13	4	26	17	20	7	20	15	15	14	12	12	12	22
10	2	4	17	14	4	26	17	20	6	18	14	14	13	12	12	11	19
10	3	0	32	24	6	37	30	35	16	43	33	33	29	27	28	24	40
10	3	1	32	24	6	37	30	34	7	34	25	24	25	20	19	19	33
10	3	4	33	26	6	37	30	35	6	33	25	23	25	20	19	19	30
10	4	0	46	36	8	47	42	47	16	55	45	44	41	37	37	33	56
10	4	1	47	36	9	48	43	49	6	48	37	34	37	29	28	29	50
10	4	4	48	37	9	49	44	50	5	48	37	34	38	29	26	29	48
10	5	0	57	47	11	55	52	58	15	65	54	52	52	45	44	41	71
10	5	1	58	48	11	55	53	58	7	59	48	44	49	38	36	39	61
10	5	4	58	47	11	56	53	59	6	58	47	43	48	37	34	37	61
20	2	0	33	25	22	34	32	34	29	55	44	45	38	39	39	33	54
20	2	1	32	25	21	33	31	33	10	39	29	29	28	24	23	21	40
20	2	4	31	24	20	33	31	33	7	34	25	23	25	20	19	19	35
20	3	0	58	48	37	55	55	58	30	74	64	64	59	57	57	50	75
20	3	1	59	47	37	54	56	58	10	62	52	49	50	41	40	40	62
20	3	4	58	48	36	54	56	59	6	60	48	43	49	38	35	38	55
20	4	0	76	66	52	69	73	75	31	86	78	77	73	71	70	64	86
20	4	1	76	65	52	70	73	75	9	78	67	62	68	57	53	57	79
20	4	4	76	65	53	70	73	76	6	76	65	59	67	55	50	57	77
20	5	0	87	80	63	80	85	87	31	93	88	86	84	81	80	77	93
20	5	1	87	79	64	80	85	86	8	88	80	74	81	71	65	72	87
20	5	4	87	80	64	81	85	86	5	88	79	73	81	70	64	72	88
50	2	0	68	57	58	64	67	68	67	93	88	88	81	83	84	76	91
50	2	1	67	57	57	63	66	67	23	77	68	66	64	60	59	55	80
50	2	4	68	58	57	63	66	67	10	71	60	56	59	50	48	49	70
50	3	0	94	90	86	90	93	93	67	99	97	97	96	96	96	92	99
50	3	1	93	88	85	88	92	92	17	95	90	87	89	84	81	83	95
50	3	4	93	88	85	89	92	93	9	94	88	83	89	82	76	82	94
50	4	0	99	97	96	97	99	99	67	100	100	99	99	99	99	98	100
50	4	1	99	97	96	97	98	99	15	99	98	96	98	95	93	96	99
50	4	4	99	97	96	97	99	99	8	99	97	95	98	94	91	95	99
50	5	0	100	100	99	100	100	100	66	100	100	100	100	100	100	100	100
50	5	1	100	100	99	99	100	100	14	100	100	99	100	99	98	99	100
50	5	4	100	100	99	100	100	100	8	100	100	99	100	99	97	99	100

* The power of GLR is taken from HAWKINS (1981), Table 1.

** The simulated level of D_1 equals .01 for $n=10$. The simulated level of D_1' equals .13 for $n=10$, .09 for $n=20$, .07 for $n=50$. The simulated level of D_2' equals .07 for $n=10$, .06 for $n=20$ and .05 or .06 for $n=50$. In all other cases the simulated levels are .05 (with sometimes .06 or .04).

Conclusions: The following observations are made from the results of Table 1.

1. The conclusions drawn by HAWKINS from the simulation study are all verified here also. Although HAWKINS wrongly takes a one-sided F-test in defining Fisher's method, it seems that in Hawkins' simulation study the correct P-values are used, based on the two-sided F-test. It is seen that the power of F given in Table 1 is very similar to the power of F, presented by HAWKINS. Among the four tests considered by

HAWKINS, namely, GLR, F, T, ψ_{F2}, it is observed that GLR is never inferior. In terms of power, the four tests can be arranged as follows:

$$\begin{aligned} &\mathrm{GLR} \geq \psi_{F2} \geq F > T \quad \text{if} \quad \tau^2 < 1 \\ &\mathrm{GLR} \geq F \geq \psi_{F2} > T \quad \text{if} \quad \tau^2 \geq 1. \end{aligned}$$

2. The power functions of F, B_3, D_1 and D_2 do not depend on τ^2. Among these four, it seems that the powers of F and D_2 are close to each other, the power of F being slightly larger. The ordering of these tests with respect to power is

$$F \overset{(\geq)}{\approx} D_2 \geq B_3 \geq D_1. \tag{10}$$

3. Since $F \geq B_3$, the combination tests involving F are better than those involving B_3. The reason for it may be that in using B_3 the structure of $\operatorname{var}(Z_{i2}) = (p+q)^{-1}(q\sigma_1^2 + p\sigma_2^2)$ is ignored.
4. The ordering of ψ_{F2}, ψ_{L2}, ψ_{T2} seems to be

$$\begin{aligned} &\psi_{F2} = \psi_{L2} \geq \psi_{T2}, \quad \text{if} \quad \tau^2 < 1 \\ &\psi_{F2} = \psi_{T2} \geq \psi_{L2}, \quad \text{if} \quad \tau^2 \geq 1. \end{aligned}$$

5. GLR seems to be better than both F and ψ_{F2}. When $\tau^2 = 0$, clearly GLR is better. As τ^2 gets larger, F has power getting closer to that of GLR. Comparison of the GLR, D_2, F and ψ_{F2} leads to the ordering

$$\begin{aligned} &\mathrm{GLR} > \psi_{F2} \geq D_2 = F \quad \text{if} \quad \tau^2 < 1 \\ &GLR \geq D_2 = F \geq \psi_{F2} \quad \text{if} \quad \tau^2 \geq 1. \end{aligned}$$

6. It is clear from Table 1, that U, which is simple to compute, has power close to that of GLR. We therefore recommend the use of U.

5 Theoretical comparisons and open questions

To investigate the performance of the tests several efficiency concepts are available. Here we consider Pitman efficiency, (approximate) Bahadur efficiency and local limits of the latter.

For Pitman efficiency limiting distributions under local (contiguous) alternatives are needed. As in section 3 alternatives of the form

$$\frac{\sigma_1^2}{\sigma_2^2} = 1 + \theta n^{-1/2} \qquad \theta \in \mathbb{R} \tag{11}$$

are considered, keeping the nuisance parameter τ^2 fixed, $0 \leq \tau^2 < \infty$. By standard methods we obtain the following.

THEOREM 1. *Under alternatives of the form* (11) *the limiting distributions (as $n \to \infty$) of the test statistics are given by ($\chi_k^2(\delta)$ denotes a chi-square distribution with k df and non-centrality δ)*

$$T \to N(\theta c_1, 1)$$
$$B_2 \to \chi_1^2(\theta^2 c_2^2)$$
$$(F-1)\sqrt{n}c_2 \to N(\theta c_2, 1)$$
$$B_3 \to \chi_2^2\left(\frac{\theta^2 pq(p+q-2)}{2(p+q)^2}\right)$$
$$D_1 \to N(\theta, c_2^{-2})$$
$$D_2 \to N(\tfrac{1}{3}\theta, (3c_2)^{-2})$$
$$U \to N(\theta(c_1^2 + c_2^2)^{1/2}, 1).$$

The limiting distributions of ψ_{F2}, ψ_{L2} and ψ_{T2} are easily derived from Theorem 1. For example

$$\psi_{F2} \to -2\log\{2\Phi(-|N(\theta c_1, 1)|)\} - 2\log P_2(\chi_1^2(\theta^2 c_2^2)).$$

However, these limiting distributions are rather complicated and lead to Pitman efficiencies depending on the level α and the prescribed power β.

The limiting distributions of ψ_{F3}, ψ_{L3} and ψ_{T3} can not be deduced directly from Theorem 1, due to the fact that under the alternatives S_{22}^Z and R_{12} are dependent. However, by using their simultaneous limiting distribution one can obtain also the limiting distributions of ψ_{F3}, ψ_{L3} and ψ_{T3}. Since these distributions are rather complicated we do not present them here.

From Theorem 1 it is easily seen that B_2, F, D_1 and D_2 have Pitman efficiency 1 w.r.t. each other, that either of them has Pitman efficiency c_2^2/c_1^2 (≥ 1 if $p \geq 2$ and $q \geq 2$, with equality iff $p = q = 2$ and $\tau^2 = 0$) w.r.t. T and Pitman efficiency $c_2^2/(c_1^2 + c_2^2) < 1$ w.r.t. to U. The Pitman efficiency of these tests w.r.t. B_3 depends on α and β (cf. Theorem 4 of ROTHE, 1981).

As a second criterion of theoretical comparison of the tests involved here, we consider (approximate) Bahadur efficiency and its local limit. This criterion is based on the (exact) slope, i.e. the rate at which P-values converge to zero under a (fixed) alternative as n increases. Specifically, let $T^{(n)}$ be a statistic based on n observations for testing $H_0: \theta \in \Theta_0$ against $H_1: \theta \in \Theta_1$, where $\Theta_0 \cap \Theta_1 = \varnothing$ and Θ_0, Θ_1 are subsets of the parameter space Θ. For simplicity assume that H_0 is rejected for large values of $T^{(n)}$ and that the null distribution of $T^{(n)}$ (or its asymptotic null distribution in case of approximate Bahadur efficiency) is the same for all $\theta \in \Theta_0$.

DEFINITION 1. The exact slope $C(\theta)$ at θ of $\{T^{(n)}\}$ with P-values $\{P^{(n)}\}$ is defined as the a.e. limit of $-(2/n)\log P^{(n)}$ as $n \to \infty$ under the alternative distribution denoted by θ. The approximate slope $C^{(a)}(\theta)$ of $T^{(n)}$ at θ is obtained from the exact slope by replacing the P-values by the approximate P-values, where in stead of the exact null distribution of $T^{(n)}$ its limiting distribution is used.

The Bahadur efficiency of $\{T_n^{(1)}\}$ w.r.t. $\{T_n^{(2)}\}$ at θ is defined by $C_1(\theta)/C_2(\theta)$, where $C_1(\theta)$ and $C_2(\theta)$ are the respective slopes. The approximate Bahadur efficiency of $\{T_n^{(1)}\}$ w.r.t. $\{T_n^{(2)}\}$ at θ is defined by $C_1^{(a)}(\theta)/C_2^{(a)}(\theta)$, where $C_1^{(a)}(\theta)$ and $C_2^{(a)}(\theta)$ are the respective approximate slopes.

The slopes of the test statistics are derived in the following theorems. When the exact null-distribution depends on $\theta \in \Theta_0$, or is difficult to manage, we compute the approximate slope, based on the asymptotic null-distribution.

THEOREM 2. *The exact slopes of the T-test, the F-test and Bartlett's tests B_2 and B_3 are given by*

(i) $$C_T = \log(1-\varrho^2),$$

where

$$\varrho^2 = \varrho^2_{Z_{i1}Z_{i2}} = [\operatorname{corr}(Z_{i1}, Z_{i2})]^2$$
$$= (\sigma_1^2 - \sigma_2^2)^2 pq / \{[p\sigma_1^2 + q\sigma_2^2 + (p+q)^2\tau^2][q\sigma_1^2 + p\sigma_2^2]\},$$

(ii) $$C_F = C_{B2} = \log\{(\bar{\sigma}_w^2)^{p+q-2} / [(\sigma_1^2)^{p-1}(\sigma_2^2)^{q-1}]\},$$

where

$$\bar{\sigma}_w^2 = \frac{(p-1)\sigma_1^2 + (q-1)\sigma_2^2}{p+q-2},$$

and

(iii) $$C_{B3} = \log\{(\sigma_w^{*2})^{p+q-1} / [(\sigma_1^2 + \sigma_2^2 - \sigma_w^2)(\sigma_2^2)^{p-1}(\sigma_2^2)^{q-1}]\},$$

where

$$\sigma_w^2 = (p\sigma_1^2 + q\sigma_2^2)/(p+q)$$

and

$$\sigma_w^{*2} = \frac{(\sigma_1^2 + \sigma_2^2 - \sigma_w^2) + (p-1)\sigma_1^2 + (q-1)\sigma_2^2}{p+q-1}.$$

PROOF. Part (i) follows from the covariance matrix given in (6) and the result of MUDHOLKAR and SUBBAIAH (1981, p. 162). The exact slope of the F-test follows from (2.13) on p. 949 of HWANG and KLOTZ (1975). The rest of part (ii) and part (iii) follow from Theorem 2 of HSIEH (1979, p. 594). □

The approximate slopes of the combination tests based on (approximate) P-values are obtained by standard methods (cf. e.g. WIEAND, 1976, p. 1004).

THEOREM 3. *The approximate slopes of the test statistics using Fisher's combination method ψ_{F2} and ψ_{F3} are*

$$C_{Fi}^{(a)} = C_T + C_{Bi} \qquad i = 2, 3$$

The approximate slope of ψ_{Li} is the same as that of ψ_{Fi}. The approximate slope of Tippett's combination test ψ_{Ti} is

$$C_{Ti}^{(a)} = \max\{C_T, C_{Bi}\} \qquad (i = 2, 3).$$ □

We now examine approximate slopes of U, D_1 and D_2 in (7), (8) and (9).

THEOREM 4. *The approximate slope of U is*

$$C_U^{(a)} = \left[\tfrac{1}{2} c_1 \log\left(\frac{1+\varrho}{1-\varrho}\right) + c_2^2\left(\frac{\sigma_1^2}{\sigma_2^2} - 1\right)\right]^2 (c_1^2 + c_2^2)^{-1}.$$

The approximate slope of D_1 is

$$C_{D1}^{(a)} = \frac{(p-1)(q-1)}{2(p+q-2)}\left(\frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2 + \sigma_2^2 - \sigma_w^2}\right)^2.$$

The approximate slope of D_2 is

$$C_{D2}^{(a)} = \frac{9(p-1)(q-1)}{2(p+q-2)}\left[\frac{(\sigma_1^2)^{1/3} - (\sigma_2^2)^{1/3}}{\sigma_1^2 + \sigma_2^2 - \sigma_w^2)^{1/3}}\right]^2.$$

PROOF. The statistic U converges under H_0 to the absolute value of a standard normal distribution. The statistics D_1 and D_2 converge under H_0 to normal distributions with mean 0 and variance $\frac{2}{9}\left(\frac{1}{p-1} + \frac{1}{q-1}\right)$ and $\frac{2}{p-1} + \frac{2}{q-1}$, respectively. Under the alternative we have

$$\hat{c}_1 \to c_1, \quad R_{12} \to \varrho, \quad S_3^2 \to \sigma_1^2, \quad S_4^2 \to \sigma_2^2 \quad \text{and} \quad S_{22}^Z \to \sigma_1^2 + \sigma_2^2 - \sigma_w^2.$$

The results now easily follow. □

Although the approximate slopes are doubtful for fixed values of the alternative (cf. GLESER, 1964, p. 1544), they are useful when we consider local limits (cf. WIEAND, 1976). Writing $\sigma_1^2 \sigma_2^{-2} = 1 + \theta$ we therefore study the behaviour of the slopes as $\theta \to 0$.

THEOREM 5. *As $\theta \to 0$*

$$C_T = \theta^2 c_1^2 + O(\theta^3)$$

$$C_F = C_{B2} = \theta^2 c_2^2 + O(\theta^3)$$

$$C_{B3} = \frac{\theta^2}{2}\,\frac{pq(p+q-2)}{(p+q)^2} + O(\theta^3)$$

$$C_U^{(a)} = \theta^2(c_1^2 + c_2^2) + O(\theta^3)$$

$$C_{D1}^{(a)} = \theta^2 c_2^2 + O(\theta^3)$$

$$C_{D2}^{(a)} = \theta^2 c_2^2 + O(\theta^3).$$

PROOF. The results follow by Taylor expansion of the slopes. □

The results of Theorem 5 lead to the following ordering

$$T < B_2 = F = \psi_{T2} = D_1 = D_2 < \psi_{L2} = \psi_{F2} = U \leq \psi_{L3} = \psi_{F3}$$

and

$$B_2 \leq B_3 < \psi_{L3} = \psi_{F3}.$$

The ordering does not quite agree with the simulation results. Especially, the performance of the tests based on B_3 is overestimated. This is due to the criterion used. Even if we consider local limits of the slopes, the critical values based on different chi-square distributions are considered as asymptotically equivalent, because $\log\{1 - F_k(x)\}$ with F_k the distribution function of χ^2 with k df behaves like $-\frac{1}{2}x$ as $x \to \infty$, irrespective the value of k. However, for $\alpha = 0.05$ the critical values for $k = 2$ and 3 are 5.99 and 7.81, respectively. So in fact there are two opposite effects: under alternatives B_3 will be larger than B_2, while the critical value for B_3 is also larger than for B_2; the first point is picked up by the criterion, and the second point is ignored. This explains the contrast between the position of B_3, ψ_{L3} and ψ_{F3} in the ordering based on slopes and their position in the ordering based on the Monte Carlo experiment.

We conclude with some remarks and suggestions.

REMARK 3. The disadvantage of tests based on (combinations) of T, B_2 and F is of course that they do not use explicitly S_{22}^Z. This is clearly seen in Theorem 5: the best local slope for tests based on T, B_2 and F is $c_1^2 + c_2^2$, which is obtained e.g. by U. In general this is smaller than

$$c_1^2 + \frac{pq(p+q-2)}{2(p+q)^2}, \tag{12}$$

which seems to be the best available local slope for tests based on all observations. Note that in the particular case $p = q$ (which is used in the simulations) $c_1^2 + c_2^2$ equals (12). Intuitively it is not quite clear why in this case it is not important (in the above sense) to forget about S_{22}^Z.

REMARK 4. Testing H_{02}^* means testing $\sigma_1^2 = \sigma_2^2$ against $\sigma_1^2 \neq \sigma_2^2$ based on S_{22}^Z, S_3^2, S_4^2, or, in other words, testing $\sigma_1^2 = \sigma_2^2$ against $\sigma_1^2 \neq \sigma_2^2$ based on

$$(p+q)^{-1}(q\sigma_1^2 + p\sigma_2^2)\chi_{n-1}^2, \quad \sigma_1^2\chi_{n(p-1)}^2, \quad \sigma_2^2\chi_{n(q-1)}^2,$$

where the three chi-square distributions are independent. This is a particular case of a so called curved exponential family (cf. EFRON, 1975). This theory may be applied to obtain a motivated suitable test.

REMARK 5. From a theoretical point of view B_2, F, D_1 and D_2 are equivalent (at first order) both in terms of Theorem 1 (and the implied Pitman efficiencies) and in terms of Theorem 5. Hence, based on Pitman or Bahadur efficiency the use of D_1 or D_2 in stead of B_2 or F does not give any gain. So, asymptotically S_{22}^Z plays only a minor role in D_1 and

D_2. Therefore, it seems not very useful to combine T and D_1 or D_2 in a similar way as leading to U in (7).

Another possibility is to combine T and B_3. Note however, that in contrast to the null case, under alternatives T and B_3 are dependent, which is caused by S_{22}^{Z}. A similar remark applies to the combination of T, F and F^*, or U and F^*.

Acknowledgment

The authors are grateful to a referee for careful reading of the manuscript and for the helpful suggestions leading in particular to Remarks 1 and 2.

APPENDIX A

In this appendix the maximum likelihood estimators $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$ and $\hat{\tau}^2$ of σ_1^2, σ_2^2 and τ^2 are derived if $p=q=1$. It is shown that under $\tau^2=0$, $\hat{\tau}^2$ is irregular in the sense that $(\hat{\tau}^2-\tau^2)\sqrt{n}$ is not asymptotically normal.

To obtain $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$ and $\hat{\tau}^2$ it is easily seen that we have to maximize

$$L(\sigma_1^2,\sigma_2^2,\tau^2)=\{(\sigma_1^2+\tau^2)(\sigma_2^2+\tau^2)(1-\varrho^2)\}^{-n/2}\times$$
$$\exp\left\{-\tfrac{1}{2}\frac{1}{1-\varrho^2}\left[\frac{nv_{11}}{\sigma_1^2+\tau^2}-\frac{2\tau^2 nv_{12}}{(\sigma_1^2+\tau^2)(\sigma_2^2+\tau^2)}+\frac{nv_{22}}{\sigma_2^2+\tau^2}\right]\right\}$$

under the conditions given by (2), where

$$\varrho=\tau^2(\sigma_1^2+\tau^2)^{-1/2}(\sigma_2^2+\tau^2)^{-1/2}$$
$$v_{rs}=n^{-1}\sum_{i=1}^{n}(x_{ir}-\bar{x}_r)(x_{is}-\bar{x}_s),\quad \bar{x}_r=n^{-1}\sum_{i=1}^{n}x_{ir},\quad r,s=1,2,$$

and x_{ir} is the realisation of X_{ir}, $i=1,\ldots,n$, $r=1,2$. Since

$$P_{\sigma_1^2,\sigma_2^2,\tau^2}(V_{11}>0, V_{22}>0, V_{11}-2V_{12}+V_{22}>0)=1$$

it is assumed w.l.o.g. that $v_{11}>0$, $v_{22}>0$ and $v_{11}-2v_{12}+v_{22}>0$.

By standard theory under the less restrictive conditions

$$\sigma_1^2+\tau^2>0,\quad \sigma_2^2+\tau^2>0,\quad -1<\varrho<1,$$

L is maximized by taking

$$\widehat{\sigma_1^2+\tau^2}=v_{11},\quad \widehat{\sigma_2^2+\tau^2}=v_{12}\quad\text{and}\quad \hat{\varrho}=v_{12}(v_{11}v_{22})^{-1/2}.$$

If the individual estimators $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$ and $\hat{\tau}^2$ are nonnegative, then of course these estimators maximize L also under (2) and are the required maximum likelihood estimators, since their values are in the parameter space. Therefore

case (i): $0\le v_{12}\le v_{11}$ and $v_{12}\le v_{22}$
$\hat{\tau}^2=v_{12},\quad \hat{\sigma}_1^2=v_{11}-v_{12},\quad \hat{\sigma}_2^2=v_{22}-v_{12}.$

Next consider the case $\nu_{12} > \nu_{11}$. Define

$$a = \sigma_2^2 d^{-1}, \quad b = \sigma_1^2 d^{-1}, \quad c = \tau^2 d^{-1}$$

with $d = (\sigma_1^2 + \tau^2)(\sigma_2^2 + \tau^2)(1 - \varrho^2)$, and

$$x = \nu_{11}, \quad y = \nu_{22} \quad \text{and} \quad z = \nu_{11} - 2\nu_{12} + \nu_{22}.$$

Then we have to maximize

$$\begin{aligned} L^*(a,b,c) &= \{L(\sigma_1^2, \sigma_2^2, \tau^2)\}^{2/n} \\ &= (ab + ac + bc) \exp\{-ax - by - cz) \end{aligned}$$

under the conditions

$$a \geq 0, \quad b \geq 0, \quad c \geq 0, \quad d > 0. \tag{13}$$

Since $\nu_{12} > \nu_{11}$ we have $y > x + z$ and hence

$$\begin{aligned} \max_{\substack{a\geq 0, b\geq 0 \\ c\geq 0, d>0}} L^*(a,b,c) &\leq \max_{\substack{a\geq 0, b\geq 0 \\ c\geq 0, d>0}} (ab + ac + bc) \exp\{-ax - b(x+z) - cz\} \\ &= \max_{\substack{a\geq 0, b\geq 0 \\ c\geq 0, d>0}} \{(a+b)(b+c) - b^2\} \exp\{-(a+b)x - (b+c)z\} \\ &\leq \max_{\substack{a\geq 0, b\geq 0 \\ c\geq 0, d>0}} (a+b)(b+c) \exp\{-(a+b)x - (b+c)z\} \\ &= \max_{a>0, c>0} L^*(a,0,c). \end{aligned}$$

Therefore, if $y > x + z$, L^* is maximized under condition (13) by taking $\hat{a} = x^{-1}$, $\hat{b} = 0$, $\hat{c} = z^{-1}$. This leads to

case (ii): $\nu_{12} > \nu_{11}$
$\hat{\tau}^2 = \nu_{11}, \quad \hat{\sigma}_1^2 = 0, \quad \hat{\sigma}_2^2 = \nu_{11} - 2\nu_{12} + \nu_{22}.$

Similarly one obtains

case (iii): $\nu_{12} > \nu_{22}$
$\hat{\tau}^2 = \nu_{22}, \quad \hat{\sigma}_1^2 = \nu_{11} - 2\nu_{12} + \nu_{22}, \quad \hat{\sigma}_2^2 = 0$

and

case (iv): $\nu_{12} < 0$
$\hat{\tau}^2 = 0, \quad \hat{\sigma}_1^2 = \nu_{11}, \quad \hat{\sigma}_2^2 = \nu_{22}.$

Now it easily follows that under $\tau^2 = 0$

$$P(\hat{\tau}^2 = \tau^2) = P(\hat{\tau}^2 = 0) = P(V_{12} < 0) = \tfrac{1}{2}$$

and thus $(\hat{\tau}^2 - \tau^2)\sqrt{n}$ is not asymptotically normal.

References

BASU, D. (1955), On statistics independent of a complete sufficient statistic, *Sankhya 15*, 377-380.

COX, D. R. and N. REID (1987), Approximations to noncentral distributions, *Canadian Journal of Statistics 15*, 105-114.

EFRON, B. (1975), Defining the curvature of a statistical problem (with applications to second order efficiency), *Annals of Statistics 3*, 1189-1242.

GLESER, L. J. (1964), On a measure of test efficiency proposed by R. R. Bahadur, *Annals of Mathematical Statistics 35*, 1537-1544.

HAWKINS, D. M. (1981), Testing for equality of variance of correlated normal variables, *Statistica Neerlandica 35*, 39-47.

HSIEH, H. K. (1979), On asymptotic optimality of likelihood ratio tests for multivariate normal distributions, *Annals of Statistics 7*, 592-598.

HWANG, T. Y. and J. H. KLOTZ (1975), Bahadur efficiency of linear rank statistics for scale alternatives, *Annals of Statistics 3*, 947-954.

JOHNSON, N. L. and S. KOTZ (1970a), *Distributions in Statistics: Continuous Univariate Distributions-1*, John Wiley and Sons, New York.

JOHNSON, N. L. and S. KOTZ (1970b), *Distributions in Statistics: Continuous Univariate Distributions-2*, John Wiley and Sons, New York.

MORGAN, P. (1939), A test for the significance of the difference between two variances in a sample from a normal bivariate population, *Biometrika 31*, 13-19.

MUDHOLKAR, G. S. and E. O. GEORGE (1979), The logit statistic for combining probabilities - An overview, in: J. S. Rustagi (ed.) (1979), *Proceedings of International Symposium on Optimization and Statistics*, Academic Press, New York, 345-365.

MUDHOLKAR, G. S. and P. SUBBAIAH (1981), Complete independence in the multivariate normal distribution, in: C. Taillie et al. (eds.) (1981), *Statistical Distributions in Scientific Work 5*, D. Reidel Publishing Company, Boston, 157-168.

PITMAN, E. J. G. (1939), A note on normal correlation, *Biometrika 31*, 9-12.

ROTHE, G. (1981), Some properties of the asymptotic relative Pitman efficiency, *Annals of Statistics 9*, 663-669.

WIEAND, H. S. (1976), A condition under which the Pitman and Bahadur approaches to efficiency coincide, *Annals of Statistics 4*, 1003-1011.

Received: February 1989, Revised: December 1991.