**Significance test in bivariate canonical correlation analysis**

By B. Nielsen
*Nuffield College, Oxford OX1 1NF, U.K.*
e-mail: bent.nielsen@nuf.ox.ac.uk
15 September 1997

Summary

Canonical correlation analysis provides an example of a significance test which is not even asymptotically similar. For most values of the nuisance parameter a standard asymptotic distribution approximation applies, whereas for some particular parameter values, which may not be of practical interest, a different approximation applies. Therefore the asymptotic distribution and also Bartlett corrections do not have a uniform effect. It is suggested to use an asymptotic approximation which depends continuously on the nuisance parameter

*Some key words:* Canonical correlations, distribution approximation, nuisance parameters, significance test.

## 1. Introduction

In canonical correlation analysis it is often of interest to test the hypothesis that some of the canonical correlations are zero. The distribution of the likelihood ratio criterion for this hypothesis depends, even asymptotically, on the value of the non-restricted canonical correlations. However, except for a subset of the parameter space of Lebesgue measure zero and also of no practical interest, a limiting $\chi^2$-distribution applies. This distribution approximation is not accurate and a scalar correction to the criterion which depends on the unknown parameters has been suggested by Lawley (1959). Even the correction has not got a uniform effect, in fact, the scaling factor might be negative although the test criterion is positive.

A different approach is applied, which is to let the nuisance parameters change with sample size in the asymptotic argument. Thereby a distribution approximation is found which depends continuously on the unknown parameters. Moreover, the second order terms in the asymptotic expansions of the moments of the criterion have a rather nice behaviour, albeit not exactly that of a Bartlett correction. This approach was suggested by Nielsen (1997) in a cointegration context. Cointegration is basically an application of canonical correlation analysis to time series. Distributional analysis for cointegration is obviously more complicated and, unlike in the present paper, no analytic arguments were given.

The exact density for sample canonical correlations was found by Constantine (1963). This expression is rather complicated and therefore the present paper is limited to the simplest bivariate case. The analysis is given primarily in terms of expansions of the expectation of the relevant likelihood test criterion.

In Section 2 the general statistical model is presented. Next, the simulated expectation of the bivariate test criterion is given in Section 3 as a benchmark for the subsequent analysis. In Section 4 Bartlett's degrees of freedom correction and next in Section 5 Lawley's correction are discussed. The new approach is explained in Section 6 by an informal analysis of Constantine's density and moment expansions are described in Section 7. The new approach involves non-standard asymptotic distributions which are discussed in Section 8. In Section 9 the new approximation is evaluated by comparison with the simulated results. The paper is completed with proofs in Section 10 and a discussion in Section 11.

## 2. The statistical model and its analysis

Consider $n+1$ independent repetitions of two normally distributed vectors, $X$, $Y$, of dimension $p$ and $q$ respectively, where $p \leq q$. The variance matrices and the joint covariance matrix are denoted $\Sigma_{XX}$, $\Sigma_{YY}$ and $\Sigma_{XY}$. The hypothesis of interest is that the covariance matrix, $\Sigma_{XY}$, has reduced rank of at most $k < p$.

The hypothesis is analysed in terms of canonical correlations analysis which was developed by Hotelling (1936) and Bartlett (1938). The squared population canonical correlations, $1 \geq \lambda_1^2 \geq \cdots \geq \lambda_p^2 \geq 0$, are the solutions of the eigenvalue problem

$$\det\left(\lambda^2\Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\right) = 0$$

and thus the hypothesis can also be formulated as $\lambda_{k+1} = 0$. The likelihood function is maximised in terms of the squared sample canonical correlations, $1 > r_1^2 > \cdots > r_p^2 > 0$, solutions of the eigenvalue problem

$$\det\left(r^2 S_{XX} - S_{XY}S_{YY}^{-1}S_{YX}\right) = 0$$

where the matrices $S_{ij}$ are sample covariance matrices. The likelihood ratio criterion for the above hypothesis against a general alternative is given by

$$LR = -n \sum_{j=1}^{k} \log \left(1 - r_j^2\right) . \tag{1}$$

The squared sample canonical correlations are invariant with respect to orthonormal transformations of the data and therefore their distribution only depends on the parameters through the population canonical correlations. The distribution of the criterion also depends on these. For $\lambda_k \neq 0$ it can be approximated by a $\chi^2$-distribution with $(p-k)(q-k)$ degrees of freedom whereas different distributions apply for $\lambda_k = 0$. This distributional approximation depends non-continuously on the nuisance parameters.

## 3. Simulated expectation for the bivariate case

In the subsequent sections various distribution approximations are considered. These are developed by expanding the moments of the criterion. The expectation serves as a benchmark.

Due to the complicated analytic theory only the criterion for the hypothesis that the smallest of the canonical correlations is zero, $k = 1$, in the bivariate case, $p = q = 2$, is analysed. Because of the invariance mentioned above it can be assumed without loss of generality that $\Sigma_{XX}$ and $\Sigma_{YY}$ are identity matrices, $\Sigma_{XY}$ is the diagonal matrix with the elements $\lambda_1, 0$, the observations have mean zero and therefore only $n$ observations are considered. Henceforth the index of the largest population canonical correlation, $\lambda_1$, is omitted.

Table 1. *Simulated expectation of the criterion.*

| | $n$ | | | |
|---|---|---|---|---|
| $\lambda$ | 16 | 64 | 256 | 1024 |
| 0·8 | 1·116 | 1·022 | 1·002 | 0·999 |
| 0·4 | 0·767 | 0·925 | 0·986 | 0·995 |
| 0·2 | 0·584 | 0·675 | 0·885 | 0·981 |
| 0·1 | 0·531 | 0·514 | 0·650 | 0·878 |
| 0·05 | 0·515 | 0·468 | 0·496 | 0·635 |
| 0·025 | 0·508 | 0·450 | 0·448 | 0·493 |

Simulated values of the expectation of the criterion are reported in Table 1 as a function of the nuisance parameter $\lambda$ versus the sample size, $n$. The simulations are based on 100,000 replications. Note, that even for rather big values of $\lambda$ the limiting approximation, one, is not very accurate.

## 4. A degrees of freedom correction

Bartlett (1938) suggested that a degrees of freedom correction of the criterion might be useful. In the bivariate case this implies that the expectation should be better approximated by $1 + 5/2n$. Table 1 shows, in accordance with Bartlett's own observation, that this correction is only helpful for very large values of $\lambda$. The approximation is an exact second order expansion of the expectation in case the nuisance parameter, $\lambda$, takes the value one. Hotelling (1936) found that the minimum sample canonical correlation then is a partial correlation and its distribution has been found by Fisher, see Muirhead (1982).

## 5. Lawley's correction

A second order expansion of the expectation, which also applies for general values of $k, p, q$, was given by Lawley (1959). Provided that $\lambda^2$ is not "small" he found that

$$E_\lambda LR = 1 + \frac{5}{2n} + \frac{1}{n}\left(1 - \frac{1}{\lambda^2}\right) + \mathrm{O}\left(n^{-2}\right) \tag{2}$$

and suggested scaling the criterion with this expression. This would actually give a Bartlett correction, such that terms of order $n^{-1}$ would be eliminated from all moments of the criterion (Lawley, 1956). Note, that for $\lambda = 1$ the degrees of freedom correction is obtained.

The condition on $\lambda^2$ is supposedly that it should be larger than $c/n$ for some unspecified constant $c$. The sample canonical correlations are the characteristic roots of a bivariate matrix and the explicit expression for these involves a square-root which can be expanded under that condition.

By computing the approximation and comparing with the entries of Table 1 the relative error of the approximation can be found. For $\lambda^2 n \geq 40$, 10, 2·5, respectively, the relative error varies between 99-100%, 95-98%, 96-105%.

For $\lambda^2 n = 1 \cdot 28$ the approximation misses the entries in the table by a factor two and for $\lambda^2 n \leq 0 \cdot 64$ the approximation is negative.

Conditional inference would be preferable to the suggested inference due to the nuisance parameter in (2). Lawley proposed simply to replace $\lambda^2$ with its maximum likelihood estimator, $r_1^2$. Glynn and Muirhead (1978) justified this by finding that the asymptotic conditional distribution of $r_2^2$ given $r_1^2$ does not depend on $\lambda$. Evaluated with this "truncated" measure the scaled criterion has expectation one up to an error of order $n^{-2}$, provided that $r_1^2$ is not too small.

## 6. A new approach

Table 1 has the feature that the figures along the diagonals from top left to bottom right vary less than those along horizontal lines. This has previously been observed in a cointegration context by Nielsen (1997). This suggests replacing the nuisance parameter, $\lambda^2$, with $\xi/n$ and then fixing $\xi$ in the asymptotic arguments. This gives a limit distribution which is a continuous function of $\xi$ and, as it turns out, a rather good approximation to the exact distribution. In the following the density of the sample canonical correlations is analysed using the new approach.

The joint density for squared sample canonical correlations was found by Constantine (1963) in terms of a hypergeometric function with matrix arguments, see also Muirhead (1982, p. 557). For general dimensions zonal polynomials are involved in the definition of these functions, however, in the present case a simplification is obtained. When the smallest population canonical correlation is zero the density for $(x, y) = n(r_1^2, r_2^2)$ is

$$\frac{1}{4}\left(1-\frac{2}{n}\right)\left(1-\frac{3}{n}\right)(xy)^{-1/2}\left\{\left(1-\frac{x}{n}\right)\left(1-\frac{y}{n}\right)\right\}^{(n-5)/2}(x-y)$$
$$\left(1-\lambda^2\right)^{n/2} {}_2F_1\left\{\begin{array}{c} n/2, n/2 \\ 1 \end{array}\middle| \begin{pmatrix} \lambda^2 & 0 \\ 0 & 0 \end{pmatrix}, \frac{1}{n}\begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix}\right\} \tag{3}$$

where $n > x > y > 0$. Now, the hypergeometric function with two matrix arguments can be rewritten as a hypergeometric function with one matrix argument which is then integrated over a Stiefel manifold with respect to an invariant measure, see Muirhead (1982, p. 67,260) or formula (2.8) of Glynn and Muirhead (1978). Since one of the two matrix arguments has rank one it follows that

$$
\begin{aligned}
& {}_2F_1\left\{ \begin{matrix} n/2, n/2 \\ 1 \end{matrix} \middle| \begin{pmatrix} \lambda^2 & 0 \\ 0 & 0 \end{pmatrix}, \frac{1}{n}\begin{pmatrix} x & 0 \\ 0 & y \end{pmatrix} \right\} \\
= \; & \frac{2}{\pi}\int_0^{\pi/2} {}_2F_1\left\{ \begin{matrix} n/2, n/2 \\ 1 \end{matrix} \middle| \frac{\lambda^2}{n}\left(x\cos^2\theta + y\sin^2\theta\right) \right\} d\theta. \qquad (4)
\end{aligned}
$$

Most terms of the expression (3) are well-behaved for large $n$, however, for $\lambda$ fixed the last two terms converge to zero respectively infinity for increasing $n$. Thus Glynn and Muirhead (1978) need to apply Laplace's method to approximate the integral in equation (4). Now, by the normalisation $\lambda^2 = \xi/n$ the first of these two terms, the $n/2$-th power of $1 - \lambda^2$, can be approximated by an exponential function and the hypergeometric function can be approximated in terms of modified Bessel functions using (4) and Hansen's confluence, see Watson (1958) and also Fields Theorem (Luke, 1969, p. 52), give

$$
{}_2F_1\left\{ \frac{n}{2}, \frac{n}{2}; 1 \middle| \left(\frac{x}{n}\right)^2 \right\} = I_0(x) + \frac{x^2}{2n} I_2(x) + \mathrm{O}\left(n^{-2}\right).
$$

For large arguments the modified Bessel function increases at an exponential order. However, it turns out that the density is actually exponentially decreasing for increasing $x$ whenever $\xi$ is much smaller than $n$. In this case a good approximation can be provided for the distribution of the criterion and this is exactly the situation which is of practical interest.

## 7. Expansion of the moments

The moments of the criterion can now be expanded using the suggested approach. Computational algorithms for the involved Bessel functions are given by Abramowitz and Stegun (1965, p. 378).

The asymptotic expectation obtained by the new approach is given by

$$
E_1(\xi) = 2 + \frac{\xi}{2} - \frac{\pi}{32}\exp(-\xi/4)\left\{(4+\xi)\, I_0\left(\frac{\xi}{8}\right) + \xi I_1\left(\frac{\xi}{8}\right)\right\}^2. \qquad (5)
$$

It follows that the expectation for $\xi = 0$ is $2 - \pi/2 \sim 0 \cdot 429$ whereas for $\xi \to \infty$ it tends to one, the expectation of the $\chi^2$-distribution. Moreover, the function (5) appears to be increasing in $\xi$.

Similarly the asymptotic second moment is given by

$$E_2(\xi) = 10 + 5\xi + \frac{\xi^2}{2} - \frac{\pi}{32}\exp(-\xi/4)\left\{\left(96 + 72\xi + 16\xi^2 + \xi^3\right) I_0^2\left(\frac{\xi}{8}\right) \right.$$
$$\left. + \left(28 + 12\xi + \xi^2\right) 2\xi I_0\left(\frac{\xi}{8}\right) I_1\left(\frac{\xi}{8}\right) + (8+\xi)\,\xi^2 I_1^2\left(\frac{\xi}{8}\right)\right\}. \quad (6)$$

This has the value $10 - 3\pi \sim 0 \cdot 575$ for $\xi = 0$ and it tends to 3 for $\xi \to \infty$.

The second order expansions of the first two moments are of the form

$$E_j(\xi)\left(1 + \frac{5}{2n}\right)^j + \frac{R_j(\xi)}{n}.$$

So apart from a remainder term, $R_j$, which depends on $\xi$, the properties of a Bartlett correction are obtained. The remainder term of the first moment has the expression

$$R_1(\xi) = \frac{\xi}{4}\left[-5 + \xi + \frac{\pi}{32}\exp(-\xi/4)\left\{\left(48 + 3\xi - 2\xi^2\right) I_0^2\left(\frac{\xi}{8}\right)\right.\right.$$
$$\left.\left. + \left(10 + 3\xi - \xi^2\right) 4 I_0\left(\frac{\xi}{8}\right) I_1\left(\frac{\xi}{8}\right) + (9 - 2\xi)\,\xi I_1^2\left(\frac{\xi}{8}\right)\right\}\right] \quad (7)$$

For $\xi = 0$ the remainder term $R_1$ is obviously zero. This is a case where a Bartlett correction apply and it also found that the remainder term for the second moment, $R_2$, has the value 0 for $\xi = 0$. Thus Bartlett's degrees of freedom correction, $1 + 5/2n$, apply in this case. For $\xi \to \infty$ the remainder terms behave similarly nice. The term $R_1$ is 1 and $R_2$ is 6. This correspond to a correction of $1 + 7/2n$, just as Lawley's result for $\lambda \to \infty$. The remainder term $R_1$ is illustrated in Figure 1. The second remainder term has a similar shape but it is shifted in such a way that a Bartlett correction does not apply for general $\xi$, for example, $R_1, R_2$ are zero at approximately 0·96 respectively 1·53. However, for $\xi <$1·19 then $R_1$ is smaller than 0·016 in absolute terms, whereas for $\xi <$1·88 then $R_2$ is smaller than 0·115. This means that for practical applications the remainder term can be ignored, either because $n$ is large or because $\xi$ is small.

[ **Figure 1** ]

## 8. Approximating the asymptotic distribution

The asymptotic distribution seems to be approximated very well by a Gamma distributed fitted to the first two moments.

For the case $\xi = 0$ an expression for the asymptotic density of the criterion can be found relatively easily. In this case, for large $n$, the joint density (3) reduces to

$$\frac{1}{4\sqrt{xy}}(x-y)\exp\{-(x+y)/2\},$$

for $n > x > y > 0$. The density for $y$ is now found by integration with respect to $x$. The integrand is exponentially decreasing for large $x$, thus the upper limit can be extended to infinity. The substitution $y(z+1) = x$, $ydz = dx$ then gives the following expression for the asymptotic density of the criterion

$$\frac{y}{4}\exp(-y)\int_0^\infty z(1+z)^{-1/2}\exp(-yz/2)\,dz.$$

The integral defines a confluent hypergeometric function, see Gradshteyn and Ryzhik (1965, 9.211.4). A final expression for the asymptotic density is obtained, using the formulae 7.2.2.2 and 7.11.2.9+37 of Prudnikov, Brychkov and Marichev (1990), as

$$\frac{1}{2}\exp(-y) + \frac{1-y}{4}\sqrt{\frac{2\pi}{y}}\exp\left(-\frac{y}{2}\right)\left\{1-\operatorname{erf}\left(\sqrt{\frac{y}{2}}\right)\right\}$$

with the corresponding distribution function

$$1-\exp(-y) + \frac{\pi}{2}\{F_1(y) - F_3(y)\} - \sqrt{\frac{y\pi}{2}}\exp\left(-\frac{y}{2}\right)\operatorname{erf}\left(\sqrt{\frac{y}{2}}\right).$$

Here $F_k(\cdot)$ is the distribution function of a $\chi^2$-distributed variable with $k$ degrees of freedom and erf is the error function, a convenient modification of the standard normal integral. A computational algorithm for erf is given by Abramowitz and Stegun (1965, p. 299).

The upper quantiles of this distribution can be approximated very well by a Gamma distribution fitted to the first two moments. The relative error is up to 9%, however, from the 60% quantile and upwards the relative error is at most 0·24%.

For general values of $\xi$ it is not so easy to find explicit expressions for the density. A different measure for the quality of a Gamma approximation is a comparison of third and fourth moment of the exact distribution and of the approximation for different values of $\xi$. The relative error was found to be at most 2·2% respectively 6·5%.

For $\xi \to \infty$ at least the first four moments of the asymptotic distribution are those of a $\chi^2$-distribution with 1 degree of freedom.

## 9. Approximating the finite sample distribution

Application of the degrees of freedom correction, $1 + 5/2n$, to the first moment given by (5) gives figures which are very close to those of Table 1. For $n = 256$, 1024 the difference is smaller than the simulation error. For $n = 64$ figures which are up to 2% too small are obtained. For $n = 16$ and $\lambda$ <0·8 values which are up to 4% too small are found. For $n = 16$ and $\lambda$ =0·8 the approximation is not too good. The value is 1·010 which is improved to 1·072 by including the remainder term $R_1$. However, this case of a small $n$ and a large value of $\lambda$ is probably not common in applications. The systematic under-correction would be improved by applying the correction $\exp(5/2n)$ rather than $1 + 5/2n$.

Similar figures would be found for the second moment. Therefore it is concluded that a rather good approximation is found.

## 10. Proofs

A proof for the results of Section 7 is sketched in the following using the ideas of Section 6. Details can be obtained from the author.

Let $f$ denote the joint density of $(x, y) = n(r_1^2, r_2^2)$ given by formula (3). The $k$-th moment of $y$ is

$$\int_0^n \int_0^x y^k f(x,y)\, dy\, dx = \int_0^n \int_0^1 x^{k+1} z^k f(x, xz)\, dz\, dx.$$

The second expression is obtained by the substitution $xz = y$, $xdz = dy$.

For large arguments the modified Bessel function, $I_\mu(z)$, is of order $\exp(z)/\sqrt{z}$, see Gradshteyn and Ryzhik (1980, 8.451.5) or Watson (1958, p. 203). Thus, for large $x$, the integrand of the triplet integral is of order

$$\left(1 - \frac{\xi}{n}\right)^{n/2} \left(1 - \frac{x}{n}\right)^{n/2} \exp\left(\sqrt{\xi x}\right) \sim \exp\left\{-\frac{1}{2}\left(\sqrt{x} - \sqrt{\xi}\right)^2\right\}$$

and the integrand is therefore exponentially small for large values of $x$ if $\xi$ is much smaller than $x$, and therefore much smaller than $n$. The upper limit for integration with respect to $x$ can therefore be extended and to the order $n^{-1}$ the moments of $y$ can be approximated by

$$\left(1-\frac{2}{n}\right)\left(1-\frac{3}{n}\right)\left(1-\frac{\xi}{n}\right)^{n/2}(k+1)!\,\mathcal{I}_3$$

where

$$\begin{aligned}
\mathcal{I}_1 &= \frac{1}{2}\int_0^\infty x^{k+1}\exp\left\{-\frac{1}{2}x\,(1+z)\right\}\left[1+\frac{1}{4n}\left\{10x\,(1+z)-x^2(1+z^2\right\}\right] \\
&\qquad \left\{I_0\left(\sqrt{xw}\right)+\frac{xw}{2n}I_2\left(\sqrt{xw}\right)\right\}dx, \\
\mathcal{I}_2 &= \frac{4}{\pi\,(k+1)!}\int_0^{\pi/2}\mathcal{I}_1 d\theta, \\
\mathcal{I}_3 &= \frac{1}{4}\int_0^1 z^{k-1/2}\,(1-z)\,\mathcal{I}_2 dz
\end{aligned}$$

and $w=\xi(\cos^2\theta+z\sin^2\theta)$.

*The first integral.* Introduce the notation $p=\xi r/2$, $q=\xi\,(1+r)\,/4$, $r=(1-z)/(1+z)$, $s=\sin^2\theta$, and the substitution $t^2=x$, $2t\,dt=dx$. Then,

$$\begin{aligned}
\mathcal{I}_1 &= \int_0^\infty t^{2k+3}\exp\left\{-\frac{t^2}{1+r}\right\}\left[1+\frac{1}{n}\left\{\frac{5t^2}{1+r}-\frac{t^4(1+r^2)}{2(1+r)^2}\right\}\right] \\
&\qquad \left\{I_0\left(t\sqrt{w}\right)+\frac{t^2w}{2n}I_2\left(t\sqrt{w}\right)\right\}dt.
\end{aligned}$$

Hankel's generalisation of Weber's first exponential integral, see Gradshteyn and Ryzhik (1965, 8.970.1) or Watson (1958, p. 393) can now be applied,

$$\begin{aligned}
&\int_0^\infty t^{\mu+2\alpha+1}\exp\left(-\frac{t^2}{1+r}\right)I_\mu\left(t\sqrt{w}\right)\,dt \\
&= \frac{\alpha!w^{\mu/2}}{2^{\mu+1}}\,(1+r)^{\alpha+1+\mu}\exp\left\{\frac{1}{2}w\,(1+r)\right\}L_\mu^\alpha\left(-\frac{1}{4}\omega\,(1+r)\right) \qquad (8)
\end{aligned}$$

for $\mu\in\mathbf{N}$. Thus $\mathcal{I}_1$ can be expressed in terms of the Laguerre polynomials $L_{k+1}^0, L_{k+2}^0, L_{k+3}^0$ and $L_{k+1}^2$. The last of these can be written as a combination of the first using a recurrence formula, see Gradshteyn and Ryzhik (1965, 8.971.4).

*The second integral.* The Laguerre polynomials mentioned above have the argument $ps-q$. By the explicit expression for the Laguerre polynomial, see Gradshteyn and Ryzhik (1965, 8.970.1), it is found that

$$L_\mu^0(ps-q) = \sum_{j=0}^{\mu} \frac{(-2s)^j}{j!} \sum_{l=j}^{\mu} \frac{r^l}{(l-j)!} \sum_{m=l}^{\mu} \binom{\mu}{m} \frac{(\xi/4)^m}{(m-l)!}.$$

The substitution $s = \sin^2\theta$, $ds = 2\sqrt{s}\sqrt{(1-s)}\,d\theta$ gives that the second integral can be rewritten as a finite sum of certain integrals. These integrals can be rewritten in terms of Beta functions and confluent hypergeometric functions, see Gradshteyn and Ryzhik (1965, 3.383.1),

$$\int_0^1 s^{\nu-1}(1-s)^{\mu-1}\exp(-ps)\,ds = B(\mu,\nu)\,{}_1F_1\left(\nu;\mu+\nu;-p\right).$$

Introduction of the functions, of $z$,

$$\begin{aligned} M_\mu &= \frac{1}{\pi}\int_0^1 s^{-1/2}(1-s)^{-1/2}\exp(-ps)L_\mu^0(ps-q)ds \\ &= \sum_{j=0}^{\mu} \frac{(-1)^j\,(2j-1)!!}{j!}\,{}_1F_1\left(j+\frac{1}{2};j+1;-\frac{\xi r}{2}\right) \\ &\quad \sum_{l=j}^{\mu} \frac{r^l}{(l-j)!} \sum_{m=l}^{\mu} \binom{\mu}{m} \frac{(\xi/4)^m}{(m-l)!} \end{aligned}$$

leads to the result

$$\begin{aligned} \mathcal{I}_2 &= (1+r)^{k+2}\exp(q)\left[M_{k+1}\left\{1+\frac{2}{n}(k+2)(k+3)\right\}\right. \\ &\quad \left.+\frac{k+2}{n}\left\{-(4k+7)\,M_{k+2}+(k+3)\left(1+\frac{1-r^2}{2}\right)M_{k+3}\right\}\right] \end{aligned}$$

*The third integral.* The second integral is a function of $z$ only through $r$ which suggests the substitution $r=(1-z)/(1+z)$, $(1+r)^2dz = -2dr$. Moreover, the confluent hypergeometric functions can be rewritten as

$${}_1F_1\left(j+1/2;j+1;z\right) = \exp\left(z/2\right)\left\{A_j(z)I_0\left(z/2\right)+B_j(z)I_1(z/2)\right\}$$

where $A_j, B_j$ are some ratios of polynomials. The relevant polynomials can be found using the identity 7.11.1.5 and the recurrence formulae 7.11.1.23-25 of Prudnikov, Brychkov and Marichev (1990). Since

$$\exp(q)\exp(-\xi r/4) = \exp(\xi/4)$$

the third integral can now be found using

$$\int_0^1 r^{\alpha-1}(1-r^2)^{\beta-1} I_\nu(-cr)\,dr$$
$$= \frac{(-c)^\nu \Gamma(\beta)\Gamma\left\{(\alpha+\nu)/2\right\}}{2^{\nu+1}\Gamma\left\{\beta+(\alpha+\nu)/2\right\}\Gamma(\nu+1)}\, {}_1F_2\left\{\begin{array}{c} (\alpha+\nu)/2 \\ \beta+(\alpha+\nu)/2, \nu+1 \end{array} \middle| \frac{c^2}{4}\right\}$$

which follows by integrating the Taylor series for the Bessel function term by term. The involved hypergeometric functions can be rewritten either as polynomials or as

$$A(c)I_0^2(c/2) + B(c)I_0(c/2)I_1(c/2) + C(c)I_1^2(c/2)$$

where $A, B, C$ are some ratios of polynomials. The relevant polynomials can be found using section 7.14.2 of Prudnikov, Brychkov and Marichev (1990).

Finally, the moments of the criterion are found by expanding the logarithm in equation (1).

## 11. Discussion

The new approach apparently gives a good approximation which is of primary interest in statistical analysis. However, a number of standard theoretical properties cannot be proved.

The immediate problem is that the nuisance parameter of the new approach, $\xi$, cannot be estimated consistently, since $n(r_1^2 - \lambda^2)$ converges in distribution. It is of interest to find the distribution of the criterion for a given value of the nuisance parameter; this is usually approximated by the Taylor expansion

$$\mathcal{D}\left(LR; \lambda^2, n\right) \sim \mathcal{D}\left(LR; r_1^2, n\right) + \frac{\partial}{\partial\lambda^2}\mathcal{D}\left(LR; \lambda^2, n\right)\Big|_{\lambda^2=r_1^2}\left(\lambda^2 - r_1^2\right)$$

The last term can be ignored because the term $\lambda^2 - r_1^2$ converges to zero. Due to the consistency problem a similar expansion cannot be performed in terms of $\xi$. However, the new approach gives an improved approximation of each term in the expansion above.

Because of the nuisance parameter a conditional analysis would be of most interest. Glynn and Muirhead (1978) find the result that the asymptotic conditional distribution of $r_2^2$ given $r_1^2$ does not depend on the nuisance parameter. This seems to fail for the new approach since the nuisance parameter, $\lambda$, and the arguments, $x, y$, for the density (3) are entwined in a rather complicated way, see equation (4).

A formal Bartlett correction fails to hold for the new approach, except for certain values of the nuisance parameter.

Preliminary simulations indicate that similar results would apply for general dimensions, $p, q, k$; however, I have not been able to generalise the proofs. For general $k$ the hypergeometric function cannot be reduced as simply as in formula (4). A more detailed discussion of the involved zonal polynomials seems therefore necessary. For $k = 1$, $p = 2$ but $q = 3$ the solution to the integral (8) is expressed in terms of a confluent hypergeometric function which cannot be reduced to a finite (Laguerre) polynomial.

## Acknowledgement

I gratefully acknowledge discussions with D. R. Cox and N. Shephard.

## References

Abramowitz, M. and Stegun I. A. (1965). *Handbook of mathematical functions.* Dover.

Bartlett, M. S. (1938). Further aspects of the theory of multiple regression. *Proceedings of the Cambridge Philosophical Society* **34**, 33-40.

Constantine, A. G. (1963). Some non-central distribution problems in multivariate analysis. *Annals of Mathematical Statistics* **34**, 1270-85.

Glynn, W. J. and R. J. Muirhead (1978). Inference in canonical correlation analysis. *Journal of multivariate analysis* **8**, 468-478.

Gradshteyn, I. S. and I. M. Ryzhik (1965). *Tables of integrals, series and products.* Academic Press.

Hotelling, H. (1936). Relations between two sets of variates. *Biometrika* **28,** 321-77.

Lawley, D. N. (1956). A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika* **43**, 295-303.

LAWLEY, D. N. (1959). Tests of significance in canonical analysis. *Biometrika* **46**, 59-66.

LUKE, Y. L. (1969). *The special functions and their approximations, volume 1.* Academic Press.

MUIRHEAD, R. J. (1982). *Aspects of multivariate statistical theory.* Wiley.

NIELSEN, B. (1997). On the distribution of tests for cointegration rank. *Discussion paper, Nuffield College.* http://www.nuff.ox.ac.uk/users/nielsen/.

PRUDNIKOV, A. P., YU. A. BRYCHKOV AND O. I. MARICHEV (1990). *Integrals and series, volume 3: More special functions.* Gordon and Breach Science Publishers.

WATSON, G. N. (1958). *A treatise on the theory of Bessel functions.* 2nd edition, Cambridge University Press.