Conditional test for rank in bivariate canonical correlation analysis

Bent Nielsen
Nuffield College, Oxford OX1 1NF
`bent.nielsen@nuf.ox.ac.uk`

26 October 1998

Summary. The likelihood ratio test for the hypothesis that the smallest of two canonical correlations is zero is non-similar. This means that the distribution of the test statistic depends on a nuisance parameter: the value of the largest canonical correlation. If this parameter were known it would be possible to construct a test with a desired size. However, in applications the nuisance parameter is usually estimated. It is therefore of interest to evaluate the distributional properties of such a test, conditional on the estimator. This distribution is found. Although, it depends on the nuisance parameter, the dependency seems to be negligible for practical purposes.

Keywords: Canonical correlations, Conditional test, Contiguous approximation, Non-similar test, Nuisance parameters.

1. Introduction

This paper concerns a test situation in which the asymptotic distribution of the test depends on a nuisance parameter. For most values of the nuisance parameter a standard asymptotic distribution applies, however, for one value a different asymptotic distribution applies. As a consequence, both asymptotic distributions are often poor approximations to the exact distribution of the test. Moreover, it is not clear whether a second order distribution approximation, such as a Bartlett correction, is well-behaved. The problem arises in connection with canonical correlation analysis, and consequently also in cointegration analysis, which essentially is an application to vector autoregressive time series, see Johansen (1995). The bivariate canonical correlation analysis provides the simplest example of the problem. It is argued that the approximation suggested by Nielsen (1997) gives a rather well-behaved solution to the problem.

Consider $n+1$ independent repetitions of two bivariate vectors, X, Y, with a joint normal distribution. The canonical correlations of Hotelling (1936) are defined as follows. Let Σ_{ij} and S_{ij} be the population and sample covariance matrices respectively. Then the squared population canonical correlations, $1 \geq \lambda_1^2 \geq \lambda_2^2 \geq 0$, are the solutions of the eigenvalue problem:

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

Correspondingly, the squared sample canonical correlations, $1 > r_1^2 > r_2^2 > 0$, solve:

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

The hypothesis of interest is that the rank of the covariance matrix is at most one or, equivalently, that the smallest population canonical correlation vanishes: $\lambda_2 = 0$. The likelihood ratio criterion, suggested by Bartlett (1938), is given by:

$$LR = -n \log\left(1 - r_2^2\right).$$

Under the hypothesis, the distribution of the criterion depends on the largest population canonical correlation, λ_1. For instance, for large n, the distribution can be approximated by a $\chi^2(1)$ distribution for $\lambda_1 \neq 0$ whereas for $\lambda_1 = 0$ a different asymptotic distribution applies. This distribution is described by Nielsen (1997).

The test is Bartlett adjustable. This means that the asymptotic χ^2 approximation can be improved by scaling with a factor which is basically a good approximation to the expectation of the criterion. The approximation to the expectation by Lawley (1959) is found by fixing λ_1 and letting n increase to infinity. A different and more accurate approach is to fix $\mu = \lambda_1^2 n$ as n increases. The effect of this approximation is much more uniform in λ_1, see Nielsen (1997). In both cases the adjustment factor depends on the nuisance parameter, λ_1. In applications this could be replaced by its maximum likelihood estimator, the largest sample canonical correlation, r_1. Therefore it seems reasonable to consider the conditional distribution of the adjusted test statistic for fixed values of r_1.

Section 2 describes the conditional rejection frequency of the Bartlett-adjusted test. It will be shown that this is close to the significance level when neither the nuisance parameter, μ, nor the estimator $r_1^2 n = x$ is too small. From a practical point of view this is promising since the test would be combined with a test of complete independence, $\lambda_1 = \lambda_2 = 0$. If such a test were rejected x would not be too small and the likelihood of a small μ would be small. Section 3 discusses theoretical justifications for the applied conditioning technique and the employed distribution approximation is characterised as contiguous. The mathematical derivations are detailed in the Appendices.

2. The conditional rejection frequency

The conditional rejection frequency for the likelihood ratio test is considered for fixed value of the estimated nuisance parameter.

The marginal distribution of the test criterion can be approximated rather well by a Bartlett correction to a $\chi^2(1)$ distribution. Using the parametrisation $\mu = \lambda_1^2 n$ the suggested critical value is therefore of the form:

$$c_{\mu,n} = c_\alpha E_{\mu,n}, \tag{1}$$

where c_α is the α quantile for the $\chi^2(1)$ distribution and $E_{\mu,n}$ is the marginal expectation of the test statistic. Simulations indicate that the marginal rejection frequency is very close to the significance level:

$$P_{\mu,n}(LR > c_{\mu,n}) \approx \alpha,$$

see Nielsen (1997). In applications μ would be estimated by $x = r_1^2 n$ and it would be desirable to have the same property for the conditional rejection frequency, $P_{\mu,n}(LR > c_x | x)$. This probability as well as an asymptotic expansion thereof is derived in the Appendix A.

Various approximations to the marginal expectation of the likelihood ratio test statistic have been suggested. For a fixed value of μ the expectation has the following asymptotic expansion for large sample size n:

$$E_\mu(LR)\exp(7/2n) + \{R_\mu(LR) - E_\mu(LR)\}/n. \tag{2}$$

where the asymptotic expectation E_μ and the remainder term R_μ can be expressed in terms of Bessel functions:

$$\begin{aligned} E_\mu(LR) &= 2 + \frac{\mu}{2} - \frac{\pi}{32}\exp(-\mu/4)\left\{(4+\mu)I_0\left(\frac{\mu}{8}\right) + \mu I_1\left(\frac{\mu}{8}\right)\right\}^2, \qquad (3) \\ R_\mu(LR) &= \frac{\mu}{4}\left[-5 + \mu + \frac{\pi}{32}\exp(-\mu/4)\left\{\left(48 + 3\mu - 2\mu^2\right)I_0^2\left(\frac{\mu}{8}\right)\right.\right. \\ &\quad \left.\left. + \left(10 + 3\mu - \mu^2\right)4I_0\left(\frac{\mu}{8}\right)I_1\left(\frac{\mu}{8}\right) + (9 - 2\mu)\mu I_1^2\left(\frac{\mu}{8}\right)\right\}\right], \end{aligned}$$

see Nielsen (1997). The asymptotic expectation E_μ takes the value $2-\pi/2$ for $\mu = 0$ and tends to 1, the expectation of the $\chi^2(1)$ distribution, for large μ.

The remainder term is 0 for $\mu = 0$ and tends to 1 for large μ. Further, it can be shown, that the expectation of the criterion has the following asymptotic expansion for large values of μ, n:

$$E_{\mu,n}(LR) \sim 1 + \frac{1}{n}\left(\frac{7}{2} - \frac{1}{\lambda_1^2}\right). \tag{4}$$

This approximation was found by Lawley (1959). A simulation study by Nielsen (1997) found that the expression (2) is very accurate for all values of μ, at least for $n \geq 32$. However, Lawley's approximation is only accurate for rather large values of μ.

The Figure 1 shows numerically computed conditional rejection frequencies as a function of μ for $x = 6, 12, 18$ and $n = 32, 64, 128$. The chosen significance level is $\alpha = 5\%$ and three tests are considered. Two of these are based on Lawley's approximation (4) and the approximation (2). The main conclusions are that the approximation (2) always performs better than Lawley's approximation. As expected, the difference is largest for small values of x. However, both approximations give under-sized un-conditional tests. Therefore, a third test is considered, which is based on the simpler approximation where the remainder term is ignored:

$$E_{\mu}(LR)\exp(5/2n). \tag{5}$$

For the considered values of x this approximation is smaller than the other two and the derived test is better behaved.

[**Figure**]

For small values of x the conditional rejection frequency is zero. The reason is that the test is based on the marginal distribution of the criterion. For large n this happens for $x < c_{\alpha,\infty}E_{x,\infty}$ or $x < 2.42$ for $\alpha = 5\%$. So, when the largest sample canonical correlation is sufficiently small the test is always accepted. However, in that case a test for the hypothesis of complete independence would also be accepted and there is no need to adjust the test in order to obtain a rejection frequency corresponding to the level. In contrast, Lawley's critical value is negative for $x < 1$, if n is large, and it actually leads to a zero acceptance rate.

For $x = 6$ a test for complete independence, $\lambda_1 = \lambda_2 = 0$ would typically be just accepted. The figure reflects that Lawley's approximation to the expectation is poor for this value of x. However, neither of the tests reach the significance level for large n. For the same reasons as discussed above the asymptotic rejection frequency is somewhat different from the chosen level. This frequency can be computed using formula (9) of the Appendix A.

For $x = 12$ a test for complete independence would be rejected, however, the value of x is not so large that Lawley's approximation (4) gives an accurate approximation. In contrast, the value $x = 18$ is so large that Lawley's approximation and (2) do equally well.

Figure 1 indicates that for $\mu = 0$ then the conditional rejection frequencies is not close to the significance level unless x is very large. An expansion for large x of the asymptotic rejection frequency shows that this problem persists even for large n, see formula (12) in Appendix B. With a significance level of $\alpha = 5\%$ the asymptotic rejection frequency is 3.7% for $x = 10$ and 4.4% for $x = 20$. However, as seen in the figure the problem in reaching the significance level only occurs for values of μ which are close to zero. In applications the problem can therefore be neglected since the likelihood of a small μ is small whenever x is not too small.

3. Discussion

The conclusion of the above conditional analysis is that size of the Bartlett corrected likelihood test is fairly accurate despite the nuisance parameter. Arguments for conditioning can be given by stretching the usual sufficiency and ancillarity concepts. Finally, the employed distribution approximation is characterised as contiguous.

If the conditioning variable, x, were sufficient for the nuisance parameter μ, then the conditional argument could be viewed as an attempt to obtain a Neyman structure, see Cox and Hinkley (1974, p. 135). In the considered statistical model the sufficient statistic is given by the sample covariance matrices S_{ij}, however, the variable x can be seen as approximately second-order sufficient for the nuisance parameter. The first step in such an argument is that the sample canonical correlations can be obtained from the sufficient statistic by an invariant reduction. The likelihood could then be defined from the joint distribution of the sample canonical correlations. Under the hypothesis this is given by equation (8) in the Appendix A. For fixed λ_1, not too small value of $\mu = \lambda_1^2 n$ and large n, the smallest, squared canonical correlation converges to zero at rate n^{-1} and the largest, squared canonical correlation is consistent at the usual rate of $n^{-1/2}$. Therefore, the linkage factor between the sample canonical correlations: $(r_1^2 - r_2^2)/\lambda_1^2$ depends only on r_1^2 up to order n^{-1}. Moreover, the involved hypergeometric function can be expanded so that the sample canonical correlations are not entwined up to the same order. The joint density therefore shows approximate second-order independence of the sample canonical correlations. In this sense, only the largest sample canonical correlation is relevant for likelihood purposes, see Muirhead (1982, pp. 565-567) and also Srivastava and Carter (1980).

The conditioning variable is also second-order locally ancillary for the parameter of interest, λ_2, in the sense of Cox (1980), at least up to an approximation. For fixed values of λ_1 and $\lambda_2^2 n$, not too small a value of μ and not too large a value of λ_2, it can be argued as above that the distribution of r_1^2 does not depend on λ_2 up to order n^{-1}.

The conditioning argument could be taken further and applied in the determination of the test statistic. However, in the situations described above, large n and not too small μ, the sample canonical correlations are approximately independent and there would therefore not be much difference between a conditional approach and the applied unconditional approach.

The approximation to the marginal distribution of the criterion, which is obtained by fixing $\mu = \lambda_1^2 n$ in the asymptotic argument, is contiguous as argued in the Appendix C. Consequently, the log likelihood ratio of two probability measures taken under the hypothesis, $\lambda_2 = 0$, is approximately normal distributed, see Roussas (1972). This property and some further properties, discovered by LeCam, would be useful for considerations of local power of tests for the sub-hypothesis, $\lambda_1 = \lambda_2 = 0$, against the hypothesis, $\lambda_2 = 0$. These results cannot be applied immediately in this context where the size of a test for the hypothesis, $\lambda_2 = 0$, against a general alternative is of interest.

Acknowledgements
Discussions with D. R. Cox and S. Johansen are gratefully acknowledged.

Appendix A. The conditional distribution

The conditional distribution of the smallest sample canonical correlation, r_2^2, given the largest, r_1^2, is found for the case where the smallest population canonical correlation, $\lambda_2 = 0$. Let $g_{\mu,n}$ denote the density of $(x, y) = n\left(r_1^2, r_2^2\right)$ where $\mu = \lambda_1^2 n$. Then the conditional of θ given x is given by:

$$P_{\mu,n}\left(y > c|\, x\right) = \int_c^x g_{\mu,n}\left(x, y\right) dy \Big/ \int_0^x g_{\mu,n}\left(x, y\right) dy. \tag{6}$$

Two results are given below: a numerically stable expression for the integrated density as well as asymptotic expressions derived for large n and a fixed value of μ.

The integrated density can be represented as:

$$\int_c^x g_{\mu,n}\left(x, y\right) dy \;\propto\; \sum_{k=0}^{\infty} \left\{\frac{(n/2)_k}{k!}\right\}^2 \left(\frac{\mu x}{4n^2}\right)^k \sum_{j=0}^{k+1} \int_{c/x}^{1} z^{j-1/2} \left(1 - \frac{xz}{n}\right)^{(n-5)/2} dz$$
$$\left\{\binom{2j}{j}\binom{2k-2j}{k-j} - \binom{2j+2}{j+1}\binom{2k-2j-2}{k-j-1}\right\}, \tag{7}$$

up to a proportionality factor depending on x, μ. It has been used that binomial coefficients with integer parameters $2n, n$ are zero whenever $n < 0$. The argument is as follows. The density $g_{\mu,n}$ can be derived from a result by Constantine (1963):

$$\frac{1}{4}\left(1 - \frac{2}{n}\right)\left(1 - \frac{3}{n}\right)(xy)^{-1/2}\left\{\left(1 - \frac{x}{n}\right)\left(1 - \frac{y}{n}\right)\right\}^{(n-5)/2}(x - y)$$
$$\left(1 - \frac{\mu}{n}\right)^{n/2} \frac{2}{\pi} \int_0^{\pi/2} {}_2F_1\left\{ \begin{matrix} n/2, n/2 \\ 1 \end{matrix} \,\middle|\, \frac{\mu}{n^2}\left(x\cos^2\theta + y\sin^2\theta\right)\right\} d\theta, \tag{8}$$

for $n > x > y > 0$, see also Muirhead (1982, p. 260, 397) or formula (2.8) of Glynn and Muirhead (1978). Expansion of the hypergeometric function in (8) gives:

$$\int_0^{\pi/2} {}_2F_1\left\{ \begin{matrix} n/2, n/2 \\ 1 \end{matrix} \,\middle|\, \frac{\mu}{n^2}\left(x\cos^2\theta + y\sin^2\theta\right)\right\} d\theta$$
$$= \sum_{k=0}^{\infty} \left\{\frac{(n/2)_k}{k!}\right\}^2 \left(\frac{\mu x}{n^2}\right)^k \sum_{j=0}^{k} \binom{k}{j}\left(\frac{y}{x}\right)^{k-j} \int_0^{\pi/2} \cos^{2j}\theta \sin^{2(k-j)}\theta d\theta$$

The trigonometric integral can by rewritten as a Beta integral using the substitution $z = \cos^2\theta$, $dz = -2\cos\theta\sin\theta d\theta$:

$$\frac{2}{\pi}\int_0^{\pi/2}\cos^{2j}\theta\sin^{2(k-j)}\theta d\theta = \frac{(2j)!\,(2k-2j)!}{j!\,(k-j)!k!4^k}$$

Using the substitution $zx = y$ for the other integral then gives:

$$\begin{aligned}\int_c^x g_{\mu,n}(x,y)\,dy \quad &\propto \quad \sum_{k=0}^{\infty}\left\{\frac{(n/2)_k}{k!}\right\}^2\left(\frac{\mu x}{4n^2}\right)^k\sum_{j=0}^{k}\binom{2j}{j}\binom{2k-2j}{k-j}\\ &\qquad \int_{c/x}^1 z^{k-j-1/2}(1-z)\left(1-\frac{xz}{n}\right)^{(n-5)/2}dz\end{aligned}$$

Change of summation index j into $k-j$ and separation of the term $(1-z)$ gives the formula (7).

Asymptotic formulae for the integrated density. For large n the hypergeometric function in the density (8) can be replaced by a Bessel function using Hansen's confluence formula, see Watson (1944, p. 154):

$${}_2F_1\left\{\begin{array}{c}n/2, n/2\\1\end{array}\middle|\left(\frac{v}{n}\right)^2\right\}\sim I_0(v).$$

The problem of finding addition formulae for hypergeometric function is now simpler. Neumann's Addition Theorem for Bessel functions, see Watson (1944, p. 358) gives:

$$\begin{aligned}&\int_0^{\pi/2} I_0\left\{\sqrt{\mu\left(x\cos^2\theta + y\sin^2\theta\right)}\right\}d\theta\\ = \quad & I_0\left\{\frac{\sqrt{\mu x}\,(1+\sqrt{z})}{2}\right\}I_0\left\{\frac{\sqrt{\mu x}\,(1-\sqrt{z})}{2}\right\}, \qquad \text{for } z = y/x,\end{aligned}$$

and consequently, for large n, the integrated density has asymptotic expansion:

$$\begin{aligned}&\int_c^x g_{\mu,n}(x,y)\,dy\\ \sim \quad & K\int_c^x\frac{x-y}{\sqrt{y}}\exp(-y/2)\int_0^{\pi/2} I_0\left\{\sqrt{\mu\left(x\cos^2\theta+y\sin^2\theta\right)}\right\}d\theta dy\\ = \quad & K\int_{c/x}^1\frac{1-z}{\sqrt{z}}\exp\left(-\frac{xz}{2}\right)I_0\left\{\frac{\sqrt{\mu x}\,(1+\sqrt{z})}{2}\right\}I_0\left\{\frac{\sqrt{\mu x}\,(1-\sqrt{z})}{2}\right\}dz\end{aligned} \tag{9}$$

where K is a constant depending on x, μ. It turns out to be rather involving to obtain (9) by a direct expansion of (7).

Expressions for the asymptotic conditional density of y given x or rather of $z = y/x$ given x can be derived from (9). For large values of μx an asymptotic expansion of the involved Bessel functions gives that the main term of the conditional density is proportional to:

$$(1-z)^{1/2} z^{-1/2} \exp(-xz/2),$$

see Watson (1944, p. 203). This corresponds to the asymptotic conditional density of r_2^2 given r_1^2 derived by Glynn and Muirhead (1978) for not too small values of λ_1^2:

$$\left(r_1^2 - r_2^2\right)^{1/2} \left(r_2^2\right)^{-1/2} \left(1 - r_2^2\right)^{(n-5)/2}. \tag{10}$$

However, for $\mu = 0$ the asymptotic conditional density is protional to:

$$(1-z) z^{-1/2} \exp(-xz/2) \tag{11}$$

It is easily seen that the asymptotic conditional density (10) is not valid when $\mu = 0$ and hence for small values of μ.

Appendix B: Asymptotic expansion of conditional rejection frequency.

It will be proved that for large n and $\mu = 0$ the conditional rejection frequency, $P_{0,\infty}(LR > c_{x,\infty}|x)$, has the expansion:

$$\alpha + \frac{4(1-\alpha)\exp(-x/2)}{x\sqrt{2\pi x}} \left\{1 - \frac{1}{x} + \frac{\exp(-c_\alpha/2)}{1-\alpha}\sqrt{\frac{c_\alpha}{2\pi}}\left(1 - \frac{5-3c_\alpha}{4x}\right)\right\}$$
$$-\frac{\exp(-c_\alpha/2)}{x}\sqrt{\frac{c_\alpha}{2\pi}}\left(1 - \frac{5-3c_\alpha}{4x} - \frac{53-6c_\alpha-c_\alpha^2}{4x^2}\right), \tag{12}$$

for large x.

The acceptance frequency is easier to expand than the rejection frequency. It is the ratio of integrals of the form:

$$\int_0^c g_{0,n}(x,y)\,dy$$

For $\mu = 0$ and large n the conditional density is given by (11) and up to proportionality the integral can therefore be approximated by:

$$\int_0^{c/x} z^{-1/2} (1-z) \exp(-xz/2)\, dz$$

That integral can be reformulated in terms of an incomplete Gamma integral by the substitution $y = xz/2$ and then using partial integration:

$$\begin{aligned} &\left(\frac{x}{2}\right)^{1/2} \int_0^{c/x} z^{-1/2} (1-z) \exp(-xz/2)\, dz \\ = &\left(1 - \frac{1}{x}\right) \gamma\left(\frac{1}{2}, \frac{c}{2}\right) + \frac{2}{x} \left(\frac{c}{2}\right)^{1/2} \exp(-c/2)\,. \end{aligned} \tag{13}$$

The denominator of the acceptance probability is found for $c = x$. Employing the asymptotic expansion for the incomplete Gamma function, see Gradshteyn and Ryzhik (1965, 8.356.3, 8.357), it follows for large x that:

$$\begin{aligned} &\left(\frac{x}{2\pi}\right)^{1/2} \int_0^{1} z^{-1/2} (1-z) \exp(-xz/2)\, dz \\ \sim &\left(1 - \frac{1}{x}\right) + \left(\frac{2}{x\pi}\right)^{1/2} \exp(-x/2) \\ &- \left(1 - \frac{1}{x}\right) \left(\frac{2}{x\pi}\right)^{1/2} \exp(-x/2) \left(1 - \frac{1}{x} + \frac{3}{x^2}\right) \\ \sim &\left(1 - \frac{1}{x}\right) \left\{1 + \left(\frac{2}{x\pi}\right)^{1/2} \frac{2}{x} \exp(-x/2) \left(1 - \frac{1}{x} + \frac{8}{x^2}\right)\right\}. \end{aligned} \tag{14}$$

The numerator of the acceptence probability is found for $c = c_{x,\infty}$. Using (1) and (3) the asymptotic critical value, $c_{x,\infty}$, is found to have asymptotic expansion $c_\alpha (1 - 1/x + 2/x^2)$ and the incomplete Gamma function in (13) is therefore:

$$\frac{1}{\sqrt{\pi}} \gamma\left(\frac{1}{2}, \frac{c_{x,\infty}}{2}\right) = \frac{1}{\sqrt{2\pi}} \int_0^{c_\alpha \left(1 - 1/x + 2/x^2\right)} y^{-1/2} \exp(-y/2)\, dy.$$

A second order Taylor expansion around c_α of the function defined by this integral gives:

$$\frac{1}{\sqrt{\pi}} \gamma\left(\frac{1}{2}, \frac{c_{x,\infty}}{2}\right) \sim 1 - \alpha - \frac{\exp(-c_\alpha/2)}{x} \sqrt{\frac{c_\alpha}{2\pi}} \left\{1 + \frac{9 + c_\alpha}{4x} + \frac{12 + c_\alpha}{x^2}\right\}.$$

Further expansion of the other terms of (13) gives:

$$\begin{aligned}
\exp\left(-c_{x,\infty}/2\right) &\sim \exp\left(-c_\alpha/2\right)\left\{1+\frac{c_\alpha}{2x}+\frac{c_\alpha}{x^2}\left(1+\frac{c_\alpha}{8}\right)+\frac{c_\alpha}{2x^3}\left(11+c_\alpha\right)\right\},\\
\sqrt{c_{x,\infty}} &\sim \sqrt{c_\alpha}\left(1-\frac{1}{2x}-\frac{9}{8x^2}-\frac{6}{x^3}\right),
\end{aligned}$$

and, consequently the numerator has expansion:

$$\begin{aligned}
&\left(\frac{x}{2\pi}\right)^{1/2}\int_0^{c_{x,\infty}/x} z^{-1/2}\left(1-z\right)\exp\left(-xz/2\right)dz\\
\sim\ &\left(1-\frac{1}{x}\right)\left\{1-\alpha+\frac{\exp\left(-c_\alpha/2\right)}{x}\sqrt{\frac{c_\alpha}{2\pi}}\left(1-\frac{5-3c_\alpha}{4x}-\frac{53-6c_\alpha-c_\alpha^2}{4x^2}\right)\right\}.
\end{aligned} \tag{15}$$

The acceptance probability is given as the ratio of (15) and (14) and the expression (12) follows.

Appendix C: Contiguity

The idea of fixing $\mu=\lambda_1^2 n$ in the asymptotic argument gives a contiguous approximation. In contrast, contiguity is not obtained when λ_1 is held fixed as n increases. These properties can be argued as follows.

Introduce the probability measure $P_{\lambda,n}$ under which the observations have zero expectation, the variance matrices, Σ_{xx}, Σ_{yy}, are identity matrices and the covariance matrix Σ_{xy} is diagonal with element λ_1 and 0. Consider the log likelihood, $\Lambda_{\lambda,n}$, of two such measures,

$$\begin{aligned}
&2\log\frac{\partial P_{\lambda,n}}{\partial P_{0,n}}\left(X_1,\ldots,X_{n+1},Y_1,\ldots,Y_{n+1}\right)\\
=\ &-\left(n+1\right)\log\left(1-\lambda_1^2\right)+\frac{2\lambda_1}{1-\lambda_1^2}\sum_{j=1}^{n+1}X_{j,1}Y_{j,1}-\frac{\lambda_1^2}{1-\lambda_1^2}\sum_{j=1}^{n+1}\left(X_{j,1}^2+Y_{j,1}^2\right).
\end{aligned}$$

The Central Limit Theorem gives that $\Lambda_{\lambda,n}$ is divergent under $P_{\lambda,n}$ as well as under $P_{0,n}$. The sequences $\left\{P_{\lambda,n}\right\}_{n\in\mathbf{N}}$ and $\left\{P_{0,n}\right\}_{n\in\mathbf{N}}$ are therefore not contiguous, see Roussas (1972, p. 11).

Let $\tilde{P}_{\mu,n}$ be the corresponding measure where λ_1^2 is replaced by μ/n. It follows that the limit distribution of the likelihood ratio $\tilde{\Lambda}_{\mu,n}=2\log(\partial\tilde{P}_{\mu,n}/\partial\tilde{P}_{0,n})$ is well-defined. It is normal, $N\left(\mu,4\mu\right)$, under $\partial\tilde{P}_{\mu,n}$ and $N\left(-\mu,4\mu\right)$ under $\partial\tilde{P}_{0,n}$. Therefore the sequences $\{\tilde{P}_{\mu,n}\}_{n\in\mathbf{N}}$ and $\{\tilde{P}_{0,n}\}_{n\in\mathbf{N}}$ are contiguous.

References

Bartlett, M. S. (1938). Further aspects of the theory of multiple regression. *Proc. Camb. Phil. Soc.,* **34**, 33-40.

Constantine, A. G. (1963). Some non-central distribution problems in multivariate analysis. *Ann. Math. Statist.,* **34**, 1270-85.

Cox, D. R. (1980). Local ancillarity. *Biometrika,* **67**, 279-86.

Cox, D. R. and Hinkley, D. V. (1974). *Theoretical statistics.* London: Chapman and Hall

Glynn, W. J. and Muirhead, R. J. (1978). Inference in canonical correlation analysis. *Jour. Mult. Anal.,* **8**, 468-78.

Gradshteyn, I. S. and Ryzhik, I. M. (1965). *Tables of Integrals, Series and Products.* New York: Academic Press.

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

Johansen, S. (1995). *Likelihood-based inference in cointegrated vector autoregressive models.* Oxford University Press.

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

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory.* New York: Wiley.

Nielsen, B. (1997). The likelihood ratio test for rank in bivariate canonical correlation analysis. *Biometrika,* forthcoming.

Roussas, G. G. (1972). *Contiguity of probability measures: some applications in statistics.* Cambridge University Press.

Srivastava, M. S. and Carter, E. M. (1980). Asymptotic expansions for hypergeometric functions. In *Multivariate analysis, vol. V* (ed. P. R. Krishnaiah), pp. 337-47. New York: North-Holland.

Watson, G. N. (1944). *A Treatise on the Theory of Bessel Functions,* 2nd ed. Cambridge University Press.

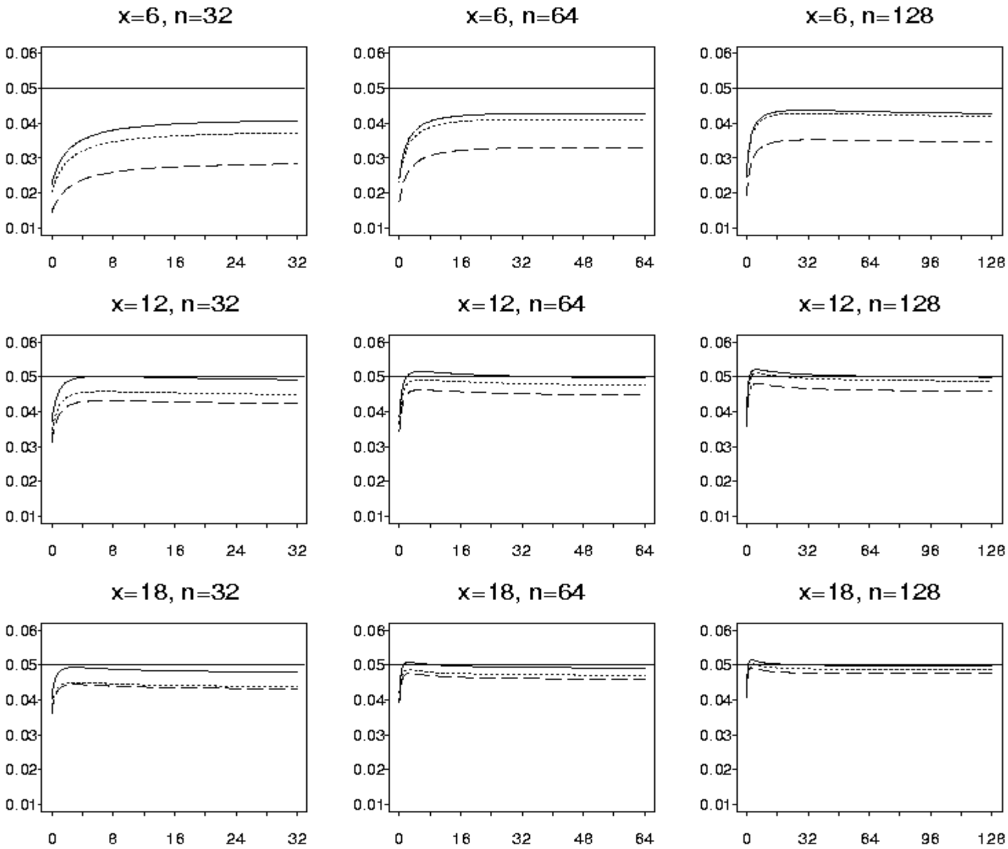

Figure 1: Rejection probabilities as function of μ. Solid line: $E_\mu \exp(5/2n)$, dotted line: second order expansion, (2), dashed line: Lawley's expression, (4).