On the distribution of tests for cointegration rank

Bent Nielsen
Nuffield College, Oxford OX1 1NF, UK
Email: bent.nielsen@nuf.ox.ac.uk
http://www.nuff.ox.ac.uk/users/nielsen/

22 May 1997, revised: 29 June 1998

Abstract: This paper analyses the likelihood test for the hypothesis of reduced cointegration rank in a Gaussian vector autoregressive model. In finite samples the rejection probability for the hypothesis may be quite different from the promised asymptotic size. An explanation is found in the fact that the test is not similar. A new asymptotic distribution which depends continuously on the nuisance parameters is suggested. This captures the functional form of the exact distribution and gives a rather good approximation. The idea is discussed for some low dimensional examples.

1. Introduction

Empirical evidence of economic equilibria can be found by cointegration analysis using the idea that a linear combination of two stochastically trending processes may be stationary. One of many tests for the number of cointegrating relations is suggested by Johansen (1988, 1995a). This test has appeal as a likelihood ratio test in a Gaussian vector autoregressive model although it is based on an asymptotic distribution approximation of the test statistic. Simulation studies of the test show that under the hypothesis the rejection probability of the test may be quite different from the promised asymptotic size. As an example consider the money demand analysis for Denmark by Johansen and Juselius (1990). One of their tests is for the hypothesis of at most one cointegrating relation. Parametric bootstrap shows that the rejection probability evaluated at the maximum likelihood estimator is around 22% when the asymptotic size is 5%.

The statistical model involves a wide range of parameters, however, the statistical analysis only has a reasonable interpretation if the first differences of the process and the cointegrating vector are stable. Although the test statistic conveniently has one asymptotic distribution under these assumptions

[0]Comments from G. Bårdsen, D.R. Cox, S. Johansen, H. Lütkepohl and A. Rahbek are gratefully acknowledged.

the test cannot even be regarded as asymptotically similar. If the assumptions are violated, different asymptotic distributions may apply. In fact, the finite sample distribution of the test statistic depends continuously on nuisance parameters and it is approximated by a step function. The asymptotic approximation is consequently poor near the steps of that function.

If the nuisance parameters are not too close to the steps it seems as though the asymptotic distribution can be improved succesfully by a Bartlett correction. For the money demand example mentioned above the size is improved from 22% to 5%. The idea of Bartlett (1937) entails scaling with the ratio of the asymptotic expectation and the finite sample expectation, which depends on the nuisance parameters. Therefore the approach is analytically rather demanding.

A different approach is to apply an asymptotic distribution which is more closely related to the exact distribution. The idea is to identify the important nuisance parameters, Θ, scale them by the sample size, $\theta = T\Theta$, and fix θ rather than Θ in the asymptotic argument. This approach has two important features. First, the statistical model and analysis is not changed so the nuisance parameter Θ is estimated by existing methods. Secondly, the approach captures the shape of the finite sample distribution and it has a nearly uniform effect in the entire parameter space. The approach will be demonstrated in this paper using a few examples. In these cases the new asymptotic distribution gives an approximation which is at least as good as the Bartlett corrected standard asymptotic distribution. While the analytic arguments for this method are also rather demanding, the benefit is that the problem is now formulated as an asymptotic problem of first order rather than of second order.

In the paper three types of cointegration models are considered. When testing for no cointegration in a first order model the test is similar and the standard asymptotic distribution gives a fine approximation. In general models the test is non-similar. This non-similarity arises from two sources. Firstly, higher order autoregressive structure allows for extra unit roots and as a consequence the test is non-similar as pointed out by for example Pantula (1989). This will be discussed in the context of a univariate second order model. Secondly, cointegration is analysed using canonical correlations. In classical canonical correlation analysis there is also a similarity problem as pointed out by Hotelling (1936). This problem is inherited by cointegration analysis. Here it will be discussed for a bivariate first order model. The new asymptotic approach is considered in connection with classical canonical correlations by Nielsen (1997c). That paper includes more analytic arguments since the observations are assumed to be independent and the problem is therefore more tractable.

A brief introduction to the general model is given in Section 2 and, next, in Section 3 the similar test for no cointegration in first order models is discussed. The univariate second order model and the bivariate first order model are considered in Section 4 and 5 respectively. Finally, Section 6 concludes with a discussion of some further issues.

2. The statistical model and the asymptotic analysis

Consider the p-dimensional, k-th order vector autoregressive model given by

$$\Delta X_t = \Pi X_{t-1} + \sum_{j=1}^{k-1} \Gamma_j \Delta X_{t-j} + \varepsilon_t, \qquad \text{for } t = 1, \ldots, T \tag{1}$$

for fixed values of $X_0, \ldots, X_{1-k}$ and independent standard Gaussian innovations with variance Ω. The parameters, $\Pi, \Gamma_1, \ldots, \Gamma_{k-1}, \Omega$ vary freely so that Ω is positive definite and symmetric. Usually deterministic components would be included in applications. Nielsen and Rahbek (1998) demonstrate how these components can be included in such a way that no further similarity problems are introduced.

The hypothesis of at most r cointegrating relations is formulated as

$$H(r) : rank\Pi \leq r < p.$$

The inequality ensures that the matrix Π can be written as the product $\alpha\beta'$, where α, β are freely varying $(p \times r)$ matrices.

The first differences of the time series and the cointegrating relation, $\beta' X_t$, can be given stationary initial distributions and therefore *interpreted* as stable phenomena under some restrictions to the characteristic polynomial of the time series,

$$A(z) = (1-z)I_p - z\Pi - \sum_{j=1}^{k-1} (1-z) z^j \Gamma_j. \tag{2}$$

The characteristic roots, solutions to the equation, $detA(z) = 0$, have to fulfil the assumptions

(A.1)	$p - rank\Pi$ roots are at one
(A.2)	the remaining roots are numerically larger than one

as proved by Johansen (1995a, Theorem 4.2, 4.7). The first assumption ensures that the cointegrating vector and the first difference of the time series do not have any unit roots, whereas the second emphasises stability rather than explosiveness or seasonal integration.

In the statistical analysis the assumptions (A.1)-(A.2) are ignored. The likelihood criterion for the hypothesis is therefore given by

$$LR\{H(r)|H(p)\} = -T \sum_{j=r+1}^{p} \log\left(1-\hat{\lambda}_j\right) \tag{3}$$

where $1 > \hat{\lambda}_1 > \cdots > \hat{\lambda}_p > 0$ are the squared sample canonical correlations of ΔX_t and X_{t-1} and both series are corrected for lagged differences, $\Delta X_{t-1}, \ldots, \Delta X_{t-k+1}$.

Since the test statistic is computed in terms of canonical correlations its distribution does not change by a non-singular linear transformation of the time series and the test is said to be invariant with respect to such transformations (Cox and Hinkley, 1974, p. 157). Thus p^2 linear restrictions can be imposed on the parameter space without altering the distribution of the test statistic. This generalises the finding that the univariate test is invariant with respect to scalar transformations and thereby not dependent on the parameter of the innovation variance.

The asymptotic distribution of the sample canonical correlations varies with the number of unit roots. Thus, the assumption (A.1) is necessary, whereas (A.2) is convenient for deriving the asymptotic distributions. The hypotheses of at most r cointegrating relations only specifies that the time series has at least $p-r$ unit roots, hence, for the asymptotic analysis it is also assumed that

$$\text{(A.3)} \qquad rank\Pi = r.$$

Under these assumptions Johansen (1995a, Theorem 6.1) proves that the r largest sample canonical correlations converge to constants and that the distribution of the likelihood criterion can be approximated by the distribution of the trace of a stochastic matrix of the form

$$\int_0^1 dW_u W_u' \left(\int_0^1 W_u W_u' du\right)^{-1} \int_0^1 W_u dW_u'. \tag{4}$$

Here W_u is a $p-r$ dimensional Brownian motion.

3. Similar tests

A test is said to be similar if it is based on a critical region K with probability

$$P_\theta K = \alpha$$

for all θ in the parameter space restricted by the hypothesis (Cox and Hinkley, 1974, p. 134). Although the property is satisfied asymptotically when the assumptions (A.1)-(A.3) are imposed the cointegration tests cannot be said to be asymptotically similar. The reason is that various different limit distributions may apply on the boundary of the parameter space given by these assumptions. However, when testing for no cointegration, $H(0)$, in a first order model, $k = 1$, then the assumptions (A.1)-(A.3) are trivially fulfilled in the entire parameter space and the test is actually exact similar. In that situation the asymptotic distribution gives a good approximation to the exact finite sample distribution of the test.

In the univariate case, $p = 1$, the distribution of the likelihood criterion converges unusually fast. With 8 observations the simulated finite sample distribution cannot be distinguished from the asymptotic distribution (Nielsen, 1997a).

For higher dimensions the convergence is obviously slower. In Table 1 the sizes of tests based on the asymptotic 95% quantile are reported for various dimensions and sample lengths. This table can be evaluated by the requirement that there should be 8-10 observations per parameter as in many applications. For the first four dimensions the relevant sample lengths are 16-20, 28-35, 40-50, 52-65. Thus, except for the univariate case the test is slightly, but not alarmingly, over-sized. Note, that compared with many classical tests this convergence is rather fast.

[**Table 1**]

A Bartlett correction works very well in this case. The idea is that the asymptotic quantiles and the exact finite sample quantiles are approximately proportional and the asymptotic distribution can therefore be improved by scaling for instance by the ratio of the asymptotic and the exact expectation of the test statistic. Bartlett (1937) argued in terms of a second order expansion of the expectation and later Lawley (1956) proved that for testing situations in case of n repeated observations the moments of the test statistic are given by

$$E\,(LR)^k = m_k\,(1 + a/n)^k + \mathrm{O}\left(n^{-2}\right) \tag{5}$$

where m_k is the k-th moment of the limiting distribution. Therefore, scaling with a second order expansion of the expecation would eliminate all second order terms from the distribution approximation.

These properties are slightly different for cointegration models. For the univariate first order model Nielsen (1997a) shows by simulation that the quantiles are close to proportional. Moreover, scaling with the ratio of the asymptotic expectation and the (simulated) exact distribution improves the distribution approximation. However, the analytic effect of the Bartlett correction is less clear. The second order term, denoted a in (5), depends on the moment which is considered. It is reduced in absolute value by a Bartlett correction but not eliminated.

Bartlett corrections based on the ratio of simulated finite sample and asymptotic expectations work fine also for the higher dimensional test of no cointegration in first order models. In that case the figures in Table 1 would be improved to be 4.9%-5.3% for $T = 25$.

It should be noted that the degrees of freedom correction suggested by Reinsel and Ahn (1992) is not very helpful. That correction approximates the ratio of exact and asymptotic expectation by $(1+pk/T)$. This is obviously not very good for the univariate case where the asymptotic distribution fits rather well. For the first four dimensions studied in Table 1 and $T = 25$ the simulated ratios are 1.008, 1.037, 1.066, 1.096 which should be compared with 1.04, 1.08, 1.12, 1.16 respectively.

4. The univariate second order model

This model provides the simplest example of a non-similar cointegration test and illustrates the autoregressive aspect of the problem. The distribution of the test depends, even asymptotically, on a nuisance parameter. In the following the statistical analysis is discussed briefly, next the asymptotic distribution of the test criterion is described, then approximations to the finite sample and finally some practical issues are considered.

4.1. The statistical analysis.

The model is reparametrisised in terms of the univariate equation

$$\Delta^2 X_t = \Pi X_{t-1} - \Gamma \Delta X_{t-1} + \varepsilon_t, \tag{6}$$

where $\Gamma = -\Gamma_1$. The likelihood ratio test criterion for the hypothesis $\Pi = 0$ is given by $-T\log(1-\hat{\lambda})$ where $\hat{\lambda}$ is the squared empirical correlation of $\Delta^2 X_t$ and X_{t-1} with both series corrected for ΔX_{t-1}.

For the distributional analysis two properties are of interest. Emperical correlations are, as mentioned above, scale invariant and the distribution of the criterion does therefore not depend on the innovation variance Ω. Further, an analysis of the characteristic polynomial for a process given by (6) reveals that the process has two unit roots if $\Gamma \neq 0$ and one otherwise. The process is I(1) if the absolute value of (1-Γ) is smaller than one.

4.2. The asymptotic distribution
The asymptotic distribution of the criterion is given as follows

Theorem 1 *(The univariate second order model)*
Let W be a standard Brownian motion and $\tilde{W}_u = \int_0^u W_v dv$.
For $\Gamma \neq 0$ the criterion converges in distribution to

$$\left\{ \left(\int_0^1 W_u dW_u \right)^2 \right\} \Big/ \int_0^1 W_u^2 du \tag{7}$$

For $\Gamma = 0$ the criterion converges in distribution to

$$\frac{\left(\int_0^1 \tilde{W}_u dW_u - \frac{\int_0^1 \tilde{W}_u W_u du \int_0^1 W_u dW_u}{\int_0^1 W_u^2 du} \right)^2}{\int_0^1 \tilde{W}_u^2 du - \frac{\left(\int_0^1 \tilde{W}_u W_u du \right)^2}{\int_0^1 W_u^2 du}} \tag{8}$$

Summary statistics of the distributions are given in Table 2.

The I(1) result, for abs(1-Γ) < 1, and the I(2) result, for $\Gamma = 0$, are well-known, see for example (4) and Johansen (1995b) respectively. The situation of a negative unit root, $\Gamma = -2$, is closely related to the work of Chan and Wei (1988), whereas the explosive case seems to have escaped the literature. The proof of Theorem 1 is given in the appendix.

A consequence of Theorem 1 is, as pointed out by Pantula (1989), that the test is not similar. It is tempting to exclude all but I(1) processes by the restriction abs(1-Γ) < 1. However, this is not of any help since the likelihood function does not necessarily have a maximum under this restriction and the pole at $\Gamma = 0$ remains a problem in finite samples. For the subsequent arguments it is convenient that there is just the one pole at $\Gamma = 0$ and that explosive roots are not a problem for the theoretical derivations.

[**Table 2**]

4.3. Approximation of the finite sample distribution

The following analysis of the small sample distribution is based on simulations. It is assumed that the initial values X_0, ΔX_0 are both zero. If this were not the case in an application a deterministic trend would be included in the model. An important feature of the distribution is the expectation. In the Table 3 simulated values of the expectation are reported for various values of Γ and T. Except for the extreme case $\Gamma = 0$ it is seen that a substantial number of observations are needed to obtain an expectation close to the asymptotic value. For $T = 24$ where the parameter to observation ratio is 1:8, and therefore better than in most cointegration applications, there is a significant distortion even for $\Gamma = 2/3$.

[**Table 3**]

One approach to approximation of the expectation is an asymptotic expansion for fixed Γ and increasing sample sizes. This has been done for $\Gamma = 1$ by Larsson (1994) and the correction seems to work in a range of Γ around $\Gamma = 1$. Larsson's idea could be pursued further.

A different approach is to fix $\gamma = \Gamma T$ in the asymptotic argument as suggested by Nielsen (1997c) in the context of canonical correlations. This corresponds to reading Table 3 diagonally from bottom-left to top-right corner. The figures indicate fast convergence, although to a continuum of new asymptotic distributions. The new asymptotic expectation can be estimated by fitting a polynomial in $1/T$. The relative error of the finite sample expectation in relation to the new asymptotic distribution now varies in the range 1.009 to 1.029 for $T = 24$, so that the relative error is largest for small Γ. These figures are very small and it follows that a good approximation to the expectation has been found. The new approach can be formalised as

Theorem 2 *(The univariate second order model)*
Let U be a univariate Ornstein-Uhlenbeck process given by the stochastic differential equation

$$dU_t = -\gamma U_t dt + dW_t$$

where W is a standard Brownian motion. Then for fixed $\gamma = \Gamma T$ the test criterion converges in distribution to a random variable of form (8) *where W is replaced by U.*

A proof can for instance be given along the line of Phillips (1988) theory for near-integrated time series.

Simulated values of the expectation and variance of the asymptotic distribution given in Theorem 2 are reported in Table 4. The distribution can be fitted well by a Γ-distribution determined from the first two moments. For example for $\gamma = 0$ and 4 the simulated asymptotic 95% quantiles are 4.90 and 4.59 as compared with Γ-fits of 4.94 and 4.61.

[**Table 4**]

The Table 5 shows the actual performance of the various available approximations for a test at 95% level for $T = 24$ and various values of Γ.

The entry "I(1)" gives simulated rejection frequencies of tests based on the asymptotic I(1) quantile of 4.129. The approximation is poor for small Γ.

Next, as "I(1) Bartlett" a Bartlett approximation is given where the critical value is scaled by the ratio of the simulated expectation and the asymptotic I(1) expectation, 1.142. This is in the spirit of the approach of Larsson (1994) although it actually works slightly better here since the exact expectation is approximated more accurately by simulation than by asymptotic expansion. It can be seen that this approach is quite good, although it works less well for the smallest values of Γ. The reason is that the quantiles of the two asymptotic distributions given in Theorem 1 are not exactly proportional and some of the proportionality underlying the Bartlett correction is lost.

The entry "new" gives rejection frequencies for tests based on the asymptotic 95% quantiles reported in Table 4. The figures are as good as those for the Bartlett correction although they are based on a first order approximation. This convergence is as fast as that seen for the similar tests discussed in the previous section.

Finally, as "new Bartlett" a Bartlett-type correction is given where the quantiles of Table 4 are scaled by the ratio of simulated expectations from Table 3 and the asymptotic expectations from Table 4. The error of this approximation is smaller than the simulation error.

[**Table 5**]

The conclusion supported by the figures in Table 5 is that a more accurate distribution approximation than the usual asymptotic I(1) distribution can be obtained analytically in one of two ways. The first approach is the Bartlett correction based on an asymptotic expansion of the expectation of the criterion for fixed value of Γ whereas the new approach is to apply the new asymptotic distribution obtained for fixed value of γ. These are respectively second order and first order asymptotic approaches which give roughly the same quality of distribution approximation.

There is an interesting analytic connection between the two approaches, which has been investigated in detail in the canonical correlation study of Nielsen (1997c). It will only be sketched for the present example, since no formal proof has been found. For fixed $\gamma = \Gamma T$ the expectation of the criterion can be expanded for large T as

$$E_{\gamma,T}\text{criterion} \approx f_0(\gamma)\left\{1 + \frac{f_1(\gamma)}{T} + \cdots\right\}. \tag{9}$$

Now, an asymptotic expansion for large γ gives

$$\begin{aligned} f_0(\gamma) &\approx 1.142\left(1 + \frac{a_1}{\gamma} + \frac{a_2}{\gamma^2} + \cdots\right) \\ f_1(\gamma) &\approx b_0 + \frac{b_1}{\gamma} + \cdots \end{aligned}$$

so that

$$\begin{aligned} E_{\gamma,T}\text{criterion} &\approx 1.142\left(1 + \frac{a_1}{\gamma} + \frac{b_0}{T} + \cdots\right) \\ &\approx 1.142\left\{1 + \frac{1}{T}\left(\frac{a_1}{\Gamma} + b_0\right) + \cdots\right\} \end{aligned} \tag{10}$$

which would be the asymptotic expansion of the criterion for not too small, fixed Γ and large T.

The Table 5 can be interpreted in terms of the above expansions. The entry "I(1)" is found using the leading term of the expansion (10). Next, the entry "Bartlett I(1)" is a simulation attempt to capture the second order term of the expansion (10). Actually, Larsson (1994) finds the second order coefficient $a_1/\Gamma + b_0$ for $\Gamma = 1$ in his analytic work. Finally, the entries "new" and "new Bartlett" are simulation based attempts to use respectively the leading term and the second order expansion of (9).

4.4.Practical implemention

In practical situations the nuisance parameter has to be estimated. Rather little is known about the effect of this. Before discussing that it is worth considering how close to zero the estimator for the nuisance parameter could be in practical situations.

Because of the similarity problem Pantula (1989) suggests first to test that the process has two unit roots, $\Pi = \Gamma = 0$. If this is rejected the hypothesis $\Pi = 0$ can be tested using the criterion above and now the I(1) distribution can be applied with more confidence. Even so the estimator for Γ could be close to zero. If, for instance, the estimator for Π is close to zero then the test for $\Pi = \Gamma = 0$ is roughly a unit root test for the first differences of the time series. Elliot, Rothenberg and Stock (1996) report the asymptotic power envelope for such tests. For $\Gamma = 7/24$, $T = 24$ the power is approximately 50% which illustrates that the parameter Γ can be quite close to zero in applications.

In practice there is an unresolved problem since the suggested distribution approximations depend on the nuisance parameter, Γ. Several estimators are available. The maximum likelihood estimators under the alternative as well as under the hypothesis could both be used and one could choose to bias-correct. Next, it has to be argued by some conditional argument that the distribution approximation is good given a fixed value of the estimator. For the simpler situation of canonical correlations such an answer is given by Glynn and Muirhead (1978), but even in that situation their argument is rather complicated.

5. The bivariate first order model

This model gives the simplest example of cointegration as a stationary linear relation of non-stationary variables. The model entails two types of similarity problems, the autoregressive aspect discussed aboved and the canonical correlation aspect discussed by Nielsen (1997c). The test for $H(1)$ is of interest here.

The model is given by (1) with $p = 2$ and $k = 1$. The hypothesis is, $H(1) : rank\Pi \leq 1$, so $r = 1$, $\Pi = \alpha\beta'$, where α,β are bivariate vectors.

For distributional analysis it is sufficient to consider processes given by the equation

$$\Delta X_t = \begin{pmatrix} A & B \\ 0 & 0 \end{pmatrix} X_{t-1} + \varepsilon_t \tag{11}$$

where $A \in \mathbf{R}$ and $B \geq 0$ and the innovations are independently standard normal distributed. This follows from the invariance of canonical correlations with respect to non-singular linear transformations. An alternative parametrisation is, for $A \neq 0$,

$$\Delta X_t = \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix} X_{t-1} + \begin{pmatrix} 1 & C \\ C & 1 \end{pmatrix}^{1/2} \varepsilon_t$$

where

$$C = \frac{B/A}{\sqrt{1 - B^2/A^2}}.$$

It can easily be seen that the assumption (A.1) is not satisfied for $A = 0, B \neq 0$, (A.2) is fulfilled for $-2 < A < 0$ and (A.3) is not satisfied for $A = B = 0$. The asymptotic distribution of the criterion is given as follows

Theorem 3 *(The bivariate first order model)*
For $A \neq 0$ the criterion converges in distribution to (7). *Moreover, the largest eigenvalue, $\hat{\lambda}_1$, converges in probability to $c/(1+c)$ for $|1 + A| < 1$ where $c = (A^2 + B^2)/(1 - (1 + A)^2)$ and one otherwise.*
For $A = 0$, $B \neq 0$ the criterion converges in distribution to (8) *and $\hat{\lambda}_1$ converges in probability to one.*
For $A = B = 0$ the normalised eigenvalues, $T\hat{\lambda}_1, T\hat{\lambda}_2$, converge jointly in distribution to the eigenvalues of

$$\int_0^1 dW_u W_u' \left(\int_0^1 W_u W_u' du \right)^{-1} \int_0^1 W_u dW_u' \tag{12}$$

where W is a bivariate standard Brownian motion. The asymptotic distribution of the criterion is given by the smallest of these eigenvalues.

Summary statistics of the distributions are given in Table 2.

For $A = B = 0$ and for $-2 < A < 0$ the result follows from Johansen (1995a, Chapter 11). Proofs for the remaining results are given in Nielsen (1997b).

The distribution of the test statistic depends on the parameters A, B in a complicated way. The following two tables show the variation of the expectation and variance, first for $B = 0$ and secondly for $A = 0$. These figures are compared with moments of asymptotic distributions obtained by the new approach. Again the idea is to fix $a = AT$ and $b = BT$ in the asymptotic argument. Accordingly the Ornstein-Uhlenbeck process given by the stochastic differential equation

$$dU_t = \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} U_t dt + dW_t \tag{13}$$

is of interest and new asymptotic distributions are obtained by replacing the Brownian motion in (12) by this process. It is seen from the tables that the new approach gives a rather good, nearly uniformly good, approximation to the finite sample moments.

[**Table 6**]
[**Table 7**]

In the Tables 4 and 7 the figures show a monotonous tendency. Similarly, the analysis of bivariate canonical correlations shows a monotonous increasing expectation, see Nielsen (1997c). This is not the case in Table 6, where $B = 0$. There is an overall minimum around $a = -3/8$ and a maximum around $a = -48$.

In application only some of the values in Table 6 and 7 are of interest. Following the sequential procedure suggested by Johansen (1995a, Chapter 12) the test would only be applied if it has been rejected that $A = B = 0$. For instance for $A = -18/24$, $B = 0$, $T = 24$ the asymptotic power is approximately 57% of a test at 5% level, see Johansen (1995a, Theorem 14.5) and it is likely that a test of no cointegration would be rejected. At this value of A a test based on the standard I(1) distribution (7) is slightly under-sized and, in particular, the degrees of freedom correction of $1 + 2/T$

suggested by Reinsel and Ahn (1992) would only make this worse, see also Kostial (1994). If $A = 0$ the asymptotic power of the test for no cointegration is approximately 62% for $B = 6/24$, $T = 24$. For this value the finite sample expectation is larger than the asymptotic expectation, even when it is scaled using the degrees of freedom correction.

6. Discussion and further suggestions

For a few examples it has been demonstrated that the asymptotic distribution of the cointegration test depends on the number of unit roots in the characteristic polynomial. Consequently, it is not always an improvement to use a degrees of freedom correction as that suggested by Reinsel and Ahn (1992). The approach developed in this paper gives a rather good approximation to the finite sample distribution: the convergence is fast and the error is nearly constant in the entire parameter space. Unfortunately the number of nuisance parameters increases dramatically by the dimension and by increased lag length.

A deterministic term can be included in the model without adding further similarity problems as discussed by Nielsen and Rahbek (1998). The previous results seem to carry over to that case and figures corresponding to Table 1 are reported by Toda (1995, Table 1).

The representation theorem of Granger, see Johansen (1995a, Th. 4.2) is the basis of the I(1) analysis and the assumptions should be considered in applications. A root near one not only distorts the finite sample distribution dramatically as demonstrated above, it also distorts the stationarity interpretation of the results. Therefore, it may be useful to impose such a root as a unit root in for instance an I(2) model as suggested by Johansen (1995b). In that case new possibilities for interpretation may arise, see for instance Juselius (1995) and Rahbek, Kongsted and Jørgensen (1998). Further, one could hope that the I(2) distributions would be less distorted when the near-unit roots are eliminated.

For more complicated situations than those discussed here the higher order representation theorem by LaCour (1998) may be helpful. There is very little analytic work on the multivariate I(1) distribution in the literature. The Laplace transform of the sufficient statistic in a first order vector autoregressive model is studied by Abadir and Larsson (1996) and Jensen and Nielsen (1995). Cointegration in continuous time for processes such as (13) and relevant weak convergence results have been studied by Stockmarr and Jacobsen (1994).

To conclude, the small sample problem in cointegration is actually an asymptotic problem: the small sample problem vanishes when the right asymptotic theory is employed.

Appendix: Proof of Theorem 1
First, under the hypothesis the time series $\Delta^2 X_t$ corrected for ΔX_{t-1} can be rewritten as

$$\Delta^2 X_t - \Delta X_{t-1} \frac{\sum_{t=1}^T \Delta^2 X_t \Delta X_{t-1}}{\sum_{t=1}^T (\Delta X_{t-1})^2} = \varepsilon_t - \Delta X_{t-1} \frac{\sum_{t=1}^T \Delta X_{t-1} \varepsilon_t}{\sum_{t=1}^T (\Delta X_{t-1})^2}.$$

As for the levels of the time series, then summation of the equation (6) under the hypothesis gives

$$\Delta X_t - \Delta X_0 = -\Gamma (X_{t-1} - X_0 + \Delta X_0) + \sum_{j=1}^t \varepsilon_j.$$

Addition of the term ΓX_t and some rearrangements imply that

$$\Gamma X_t = S_t - (1 - \Gamma) \Delta X_t$$

with

$$S_t = \sum_{s=0}^t \varepsilon_t, \qquad \varepsilon_0 = \Gamma X_0 + (1 - \Gamma) \Delta X_0.$$

Then, for $\Gamma \neq 0$, X_{t-1} corrected for ΔX_{t-1} can be rewritten as

$$X_{t-1} - \Delta X_{t-1} \frac{\sum_{t=1}^T X_{t-1} \Delta X_{t-1}}{\sum_{t=1}^T (\Delta X_{t-1})^2} = \frac{1}{\Gamma} \left(S_{t-1} - \Delta X_{t-1} \frac{\sum_{t=1}^T S_{t-1} \Delta X_{t-1}}{\sum_{t=1}^T (\Delta X_{t-1})^2} \right).$$

As a consequence of the results above the empirical correlation, $\hat{\lambda}$, equals

$$\frac{\left[\sum_{t=1}^T S_{t-1} \varepsilon_t - \frac{\sum_{t=1}^T S_{t-1} \Delta X_{t-1} \sum_{t=1}^T \Delta X_{t-1} \varepsilon_t}{\sum_{t=1}^T (\Delta X_{t-1})^2} \right]^2}{\left[\sum_{t=1}^T S_{t-1}^2 - \frac{\left(\sum_{t=1}^T S_{t-1} \Delta X_{t-1} \right)^2}{\sum_{t=1}^T (\Delta X_{t-1})^2} \right] \left[\sum_{t=1}^T \varepsilon_t^2 - \frac{\left(\sum_{t=1}^T \Delta X_{t-1} \varepsilon_t \right)^2}{\sum_{t=1}^T (\Delta X_{t-1})^2} \right]}$$

It remains to conclude that this term equals

$$\frac{\sum_{t=2}^T S_{t-1} \varepsilon_t}{\sum_{t=1}^T \varepsilon_t^2 \sum_{t=2}^T S_{t-1}^2} + O_P \left(\frac{1}{T^2} \right)$$

in which case the Theorem follows. The initial term, ε_0, is asymptotically ignorable and thus it suffices that

$$\frac{\left(\sum_{t=1}^T \Delta X_{t-1} \varepsilon_t \right)^2}{\sum_{t=1}^T (\Delta X_{t-1})^2}$$

converges in distribution, which follows from White (1958) for any Γ, and that

$$\frac{\left(\sum_{t=2}^{T} S_{t-1}\Delta X_{t-1}\right)^2}{T^2\sum_{t=1}^{T}\left(\Delta X_{t-1}\right)^2} \tag{14}$$

converges in probability to zero. Now, (14) follows if

$$\left(\frac{1}{Tg}\sum_{t=2}^{T} S_{t-1}\Delta X_{t-1}, \frac{1}{g^2}\sum_{t=1}^{T}\left(\Delta X_{t-1}\right)^2\right) \overset{\mathcal{D}}{\to} (0, V) \tag{15}$$

where the random variable $V \neq 0$ almost surely and g is some normalisation constant. In case $|1-\Gamma| < 1$ then $\Delta X_t, \Delta S_t$ are stationary linear processes and (15) follows from Johansen (1995a, Theorem B.13) with $g = \sqrt{T}$. For the case $\Gamma = 2$ Chan and Wei (1988, Th. 2.2) show that S_t and ΔX_{t-1}, normalised by $\sqrt{T}$, converges to independent Brownian motions and the Theorems 2.4 and 3.4.1 of the same authors imply (15) with $g = T$. Finally, for the case $|1-\Gamma| > 1$ it follows from White (1958) that

$$\frac{1}{(1-\Gamma)^{2T}}\sum_{t=1}^{T}\left(\Delta X_{t-1}\right)^2$$

is convergent in distribution which gives the relevant normalisation, $g = (1-\Gamma)^T$. Further,

$$\begin{aligned}\sum_{t=2}^{T} S_{t-1}\Delta X_{t-1} &= \sum_{t=2}^{T}\sum_{j=0}^{t-1}\varepsilon_j\sum_{k=1}^{t-1}(1-\Gamma)^{t-1-k}\varepsilon_k \\ &= \sum_{j=0}^{T-1}\sum_{k=1}^{T-1}\varepsilon_j\varepsilon_k\sum_{t=\max(j-k,0)}^{T-k-1}(1-\Gamma)^t.\end{aligned}$$

By Chebychev's inequality it can be shown that this converges in probability to zero when normalised by $(1-\Gamma)^T\sqrt{T}$. ■

References

Abadir, K.M.,1995, The limiting distribution of the t ratio under a unit root, Econometric Theory 11, 775-793.

Abadir, K.M. and R. Larsson, 1996, The joint moment generating function of quadratic forms in multivariate autoregressive series, Econometric Theory 12, 682-704.

Bartlett, M.S., 1937, Properties of sufficiency and statistical tests, Proceedings of the Royal Society (London) A 160, 268-282.

Chan, N.H. and C.Z. Wei, 1988, Limiting distributions of least squares estimates of unstable autoregressive processes. Annals of Statistics 16, 367-401.

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

Elliot, G., T.J. Rothenberg and J.H. Stock, 1996, Efficient tests for an autoregressive unit root. Econometrica 64, 813-836.

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

Hotelling, H., 1936, Relations between two sets of variables, Biometrika 28, 321-377.

Jensen, S.T. and B. Nielsen, 1995, The joint Laplace transform of a quadratic function and a non-symmetric function of a Brownian motion, University of Copenhagen, Institute of Mathematical Statistics, Discussion paper 7.

Johansen, S., 1988, Statistical analysis of cointegration vectors. Journal of Economic Dynamics and Control 12, 231-54.

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

Johansen, S., 1995b, A statistical analysis of cointegration for I(2) variables. Econometric Theory 11, 25-59.

Johansen, S. and K. Juselius, 1990, Maximum likelihod estimation and inference on cointegration - with applications to the demand for money, Oxford Bulletin of Economics and Statistics 52, 160- 210.

Juselius, K., 1995, Do purchasing power parity and uncovered interest rate parity hold in the long run? An example of likelihood inference in a multivariate time-series model, Journal of Econometrics 69, 211-240.

Kostial, K., 1994, The role of the signal-noise ratio in cointegrated systems, working paper, European University Institute.

LaCour, L., 1998, A parametric characterization of integrated vector autoregressive (VAR) processes, *forthcoming* Econometric Theory 14.

Larsson, R., 1994, Bartlett corrections for unit root test statistics. Preprint of Institute of Mathematical Statistics, University of Copenhagen.

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

MacKinnon, J.G., A.A. Haug and L. Michelis, 1996, Numerical distribution functions of likelihood ratio tests for cointegration, Mimeo.

Monsour, M.J. and P.W. Mikulski, 1994, Moments of the limiting distribution for the boundary case in the first order autoregressive process, American journal of mathematical and management sciences 14, 327-347.

Nielsen, B., 1997a, Bartlett correction of the unit root test in autoregressive models, Biometrika 84, 500-504.

Nielsen, B., 1997b, Asymptotic results for cointegration tests in non-stable cases, discussion paper, Nuffield College:
http://www.nuff.ox.ac.uk/Users/Nielsen.

Nielsen, B., 1997c, The likelihood ratio test for rank in bivariate canonical correlation analysis, Biometrika, to appear.

Nielsen, B. and A. Rahbek, 1998, Similarity issues in cointegration analysis, discussion paper, Nuffield College:
http://www.nuff.ox.ac.uk/Users/Nielsen.

Pantula, S.G., 1989, Testing for unit roots in time series data, Econometric Theory 5, 256-271.

Phillips, P.C.B., 1988, Regression theory for near-integrated time series, Econometrica 56, 1021-1043.

Rahbek, A., H.C. Kongsted and C. Jørgensen, 1998, trend-stationarity in the I(2) cointegration model, to appear in Journal of Econometrics.

Reinsel, G.C. and S.K. Ahn, 1992, Vector autoregressive models with unit roots and reduced rank structure, estimation, likelihoods ratio test, and forecasting, Journal of Time Series Analysis 13, 353-375.

Stockmarr, A. and M. Jacobsen, 1994, Gaussian diffusions and autoregressive processes: Weak convergence and statistical inference, Scandinavian Journal of Statistics 21, 403-419.

Toda, H.Y., 1995, Finite sample performance of likelihood ratio tests for cointegrating rank in vector autoregressions, Econometric Theory 11, 1015-32

White, J. S., 1958, The limiting distribution of the serial correlation coefficient in the explosive case, Annals of Mathematical Statistics 29, 1188-1197.

p versus T	25	50	100	200	400
1	5.0 %	5.0 %	5.1 %	5.0 %	5.0 %
2	6.2 %	5.3 %	5.2 %	5.2 %	5.2 %
3	8.0 %	6.2 %	5.5 %	5.2 %	5.0 %
4	11.6 %	7.5 %	6.0 %	5.5 %	5.2 %

Table 1. Simulated size of test for no cointegration in a first order model.
Based on the asymptotic 95% of MacKinnon, Haug and Michelis (1996)
and 10^5 repetitions.

Section 4	Section 5	Mean	Variance	90%	95%
irrelevant	$A = B = 0$	0.671	0.714	1.73	2.37
$\Gamma \neq 0$	$A \neq 0$	1.142	2.221	2.98	4.13
$\Gamma = 0$	$A = 0,\ B \neq 0$	1.402	3.097	3.60	4.94

Table 2: Summary statistics for limit distribution.
The figures are obtained analytically for $A \neq 0$,
see Abadir (1995), Monsour and Mikulski (1994) and Nielsen (1997a)
and otherwise by simulation of moments and Γ-fit of quantiles.

Γ versus T	6	12	24	48	96	192
0	1.723	1.491	1.431	1.414	1.407	1.402
1/24	1.727	1.498	1.434	1.391	1.322	1.253
1/12	1.733	1.495	1.404	1.328	1.254	1.203
1/6	1.731	1.467	1.342	1.258	1.208	1.180
1/3	1.706	1.393	1.271	1.210	1.179	1.162
2/3	1.607	1.295	1.215	1.178	1.160	1.150
4/3	1.440	1.230	1.183	1.162	1.153	1.149

Table 3: Simulated expectation of likelihood ratio test statistic
for fixed Γ, based on 10^6 repetitions.
The asymptotic value is 1.142 for $\Gamma > 0$ and 1.402 for $\Gamma = 0$.

γ	0	1	2	4	8	16	32	∞
expectation	1.402	1.394	1.371	1.316	1.251	1.205	1.172	1.142
variance	3.097	3.000	2.867	2.673	2.491	2.386	2.299	2.221
Γ-fit of 90% quantile	3.60	3.56	3.49	3.36	3.22	3.12	3.05	2.978
Γ-fit of 95% quantile	4.94	4.88	4.78	4.61	4.43	4.31	4.22	4.129

Table 4: Asymptotic mean and variance for γ fixed.
The values for finite γ are obtained by fitting polynomial in $1/T$ to the figures in Table 3. The values for $\gamma = \infty$ are moments obtained for $\Gamma \neq 0$ in the usual asymptotic approach, see Table 2.

Γ	0	1/24	1/12	1/6	1/3	2/3	4/3
$\gamma = \Gamma T$	0	1	2	4	8	16	32
I(1)	.082	.080	.076	.068	.061	.056	.053
I(1) Bartlett	.047	.046	.045	.045	.047	.049	.049
new	.054	.054	.053	.052	.051	.050	.051
new Bartlett	.051	.050	.050	.049	.050	.050	.050

Table 5: The rejection frequency using various distribution approximations, based on 10^6 repetitions.

A	0	-1/32	-1/16	-1/8	-1/4	-1/2	-1
$a = AT$	0	-3/4	-3/2	-3	-6	-12	-24
Finite sample mean	0.698	0.694	0.735	0.837	0.975	1.097	1.132
Finite sample var.	0.769	0.732	0.771	0.914	1.235	1.770	2.109
New asymp. mean	0.670	0.670	0.718	0.825	0.972	1.103	1.155
New asymp. var.	0.707	0.682	0.737	0.886	1.212	1.735	2.160

Table 6: Simulated finite sample expecation for $T = 24$ and asymptotic expectation and variance for $a = AT$ fixed. $B = 0$ in both cases. Based on 10^6 repetitions and $T = 1536$ in the asymptotic case.

B	0	1/32	1/16	1/8	1/4	1/2
$b = BT$	0	3/4	3/2	3	6	12
Finite sample mean	0.698	0.739	0.841	1.066	1.297	1.407
Finite sample var.	0.769	0.859	1.096	1.703	2.531	3.125
New asymp. mean	0.670	0.714	0.821	1.053	1.284	1.383
New asymp. var.	0.707	0.805	1.054	1.674	2.522	3.065

Table 7: Simulated finite sample expecation for $T = 24$ and asymptotic expectation and variance for $b = BT$ fixed. $A = 0$ in both cases. Based on 10^6 repetitions and $T = 1536$ in the asymptotic case.