Econometrica, Vol. 70, No. 5 (September, 2002), 1929–1961

A SMALL SAMPLE CORRECTION FOR THE TEST OF COINTEGRATING RANK IN THE VECTOR AUTOREGRESSIVE MODEL

BY SØREN JOHANSEN[1]

With the cointegration formulation of economic long-run relations the test for cointegrating rank has become a useful econometric tool. The limit distribution of the test is often a poor approximation to the finite sample distribution and it is therefore relevant to derive an approximation to the expectation of the likelihood ratio test for cointegration in the vector autoregressive model in order to improve the finite sample properties. The correction factor depends on moments of functions of the random walk, which are tabulated by simulation, and functions of the parameters, which are estimated. From this approximation we propose a correction factor with the purpose of improving the small sample performance of the test. The correction is found explicitly in a number of simple models and its usefulness is illustrated by some simulation experiments.

KEYWORDS: Bartlett correction, small sample properties, cointegration, rank determination, trace statistic.

1. INTRODUCTION AND SUMMARY

THE FORMULATION OF LONG-RUN economic relations as cointegrating relations (see Engle and Granger (1987)) has lead to a widespread application of the vector autoregressive model for analyzing economic data. The most frequently used model is

$$\mathcal{M}_1 : \Delta X_t = \Pi X_{t-1} + \Upsilon t^{n_d} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \sum_{i=0}^{n_d-1} \Phi_i t^i + \varepsilon_t \quad (t = 1, \dots, T). \tag{1}$$

If X_t is $I(1)$ and $\Pi = \alpha\beta'$, then $\beta' X_t - E[\beta' X_t]$ is stationary, and X_t is said to cointegrate with cointegrating vector β.

Under the assumption of Gaussian errors, the derivation of the likelihood ratio (LR) test for the hypothesis $\Pi = \alpha\beta'$ and $\Upsilon = \alpha\rho'$, the so-called trace test, and the estimation of the parameters are performed by the technique of reduced rank regression (see Anderson (1951)). The asymptotic distribution of $-2\log LR$, under the assumption of i.i.d. errors is derived by Johansen (1988, 1996) and

[1] Discussions with Bent Nielsen and Henrik Hansen have been very useful for the results of this paper, and Henrik Hansen and Stefano Fachin have helped me with the calculations involved in the approximation of the moments using RATS and GAUSS. The results are available on http://www.math.ku.dk/~sjo/.

The referees have done an extremely good job with the first version of the paper. One of them even found a mistake in the formulae, a fact for which I am very grateful. Their comments greatly improved the presentation of the results.

Ahn and Reinsel (1990). It is given by a nonstandard distribution expressed in terms of a Brownian motion and the deterministic terms. It does not contain any parameters and is tabulated by simulation as a function of $n-r$ and n_d because it is analytically intractable.

There are many studies that show by simulation that the small sample properties of the trace test are different from the asymptotic properties; see for example Cheung and Lai (1993), Toda (1995), Haug (1996), and Gonzalo and Pitarakis (1999). Ahn and Reinsel (1990) and Reimers (1992) proposed a small sample correction based on degrees of freedom, which is now commonly used, and Hansen and Rahbek (2002) employ ideas of profile likelihood to derive a correction of the Dickey Fuller test.

The distribution of the likelihood ratio test statistic depends on T and θ, where θ denotes the parameters under the null hypothesis. For $T \to \infty$ the dependence on θ disappears, but not uniformly in θ. If θ is close to the boundary where the cointegrating properties change, the approximation can be very poor; see Nielsen (1997b) for a discussion.

The trace test is widely used for making inference on cointegrating rank in many econometric software packages. It is therefore of utmost importance to have a reliable procedure, which improves the asymptotic results. The obvious alternative at the moment is to check the inference by simulating the exact distribution, using for instance the estimated parameter values as representing the DGP, and generating i.i.d. Gaussian errors or possibly resampling the errors from the fitted model. This gives the same limit distribution but not necessarily an improved approximation because of the nonuniformity in the convergence.

In this paper we suggest a correction factor to the likelihood ratio test that improves the finite sample properties. The idea is that of the Bartlett correction; see Bartlett (1937). Bartlett suggested finding the expectation of the likelihood ratio test statistic and thereby correcting it to have the same mean as the limit distribution.

A more precise formulation is as follows. We let θ denote the parameters of (1) under the assumption that $\Pi = \alpha\beta'$ and $\Upsilon = \alpha\rho'$. We want to derive an approximation to $E_\theta[-2\log LR\{\Pi = \alpha\beta',\ \Upsilon = \alpha\rho' | \mathcal{M}_1\}]$, which is a function of θ and T under the assumption of Gaussian errors. The result is expressed in terms of the LR test of $\Pi^* = 0$, $\Upsilon^* = 0$ in the model for X_t^*, of dimension $n-r$,

$$(2) \qquad \mathcal{M}_1^0 : \Delta X_t^* = \Pi^* X_{t-1}^* + \Upsilon^* t^{n_d} + \sum_{i=0}^{n_d-1} \Phi_i^* t^i + \varepsilon_t^*.$$

It turns out that if $\Pi^* = 0$, $\Upsilon^* = 0$, then

$$(3) \qquad f(T, n-r, n_d) = E[-2\log LR\{\Pi^* = 0,\ \Upsilon^* = 0 | \mathcal{M}_1^0\}]$$

only depends on T, $n-r$, and n_d and can therefore be tabulated by simulation. We can find an approximation to $E_\theta[-2\log LR\{\Pi = \alpha\beta',\ \Upsilon = \alpha\rho' | \mathcal{M}_1\}]$ of the form

$$f(T, n-r, n_d)(1 + T^{-1} b(\theta)),$$

and therefore suggest use of

$$\text{(4)} \qquad \frac{f(n-r,n_d)}{f(T,n-r,n_d)}\frac{-2\log LR}{(1+T^{-1}b(\hat{\theta}))} = \frac{-2\log LR}{a(T,n-r,n_d)(1+T^{-1}b(\hat{\theta}))},$$

with $f(n-r,n_d) = \lim_{T\to\infty} f(T,n-r,n_d)$ and $a(T,n-r,n_d) = f(T,n-r,n_d)/f(n-r,n_d)$. In this paper we derive an analytic expression for $b(\theta)$. The calculation of $a(T,n-r,n_d)$ is difficult; see Larsson (1997, 1998), Nielsen (1997a), Abadir, Hadri, and Tzavalis (1999), and Doornik, Nielsen, and Rothenberg (2002), so we suggest tabulating it by simulation as is done with the limit distribution. We then propose a simplified version of the correction factor and check its usefulness by simulation.

In many situations in classical statistics with i.i.d. observations, the Bartlett correction gives a remarkable improvement of the fit; see Bartlett (1937) and Lawley (1956), who prove that the approximations of higher order cumulants are improved in the same way as the mean. In the unit root case, however, we know that we cannot expect the same. In fact Jensen and Wood (1997) show that the Dickey-Fuller test in a univariate situation cannot be Bartlett corrected, even though Nielsen (1997a) shows that one in practice can get a better fit. This idea is followed up by Bravo (1999) who shows that for the univariate Dickey Fuller test it holds that for the cumulants of order 2, 3, and 4 the coefficient of the T^{-1} term is decreased by correcting the mean. In Jacobson and Larsson (1999) a correction factor is derived for a bivariate cointegration model for a single equation cointegrating test. Again the null hypothesis is that there is no cointegration, so that under the null hypothesis no parameters are present.

It should be noted that the calculation of the correction factor (4) is tied to the particular model (1), using the idealized assumptions that the errors are i.i.d. Gaussian, the lag length and cointegrating rank correctly specified, that is, that the data generating process is contained in the statistical model. Thus, when applying the correction, it is important to check the assumptions of the model carefully. We believe that the calculation is useful as a complement to the asymptotic analysis since it analyses the effect of the nuisance parameters and demonstrates that an uncritical use of asymptotic tables can be misleading.

In the following we repeatedly find stochastic expansions of the form

$$Q_T = A_0(T) + T^{-1/2}A_1(T) + T^{-1}A_2(T) + \cdots,$$

where $A_i(T) \in O_P(1)$. We write

$$Q_T \overset{1}{=} A_0(T) + T^{-1/2}A_1(T) + T^{-1}A_2(T)$$

to indicate that we have kept terms of order T^{-1}, and similarly, $Q_T \overset{0}{=} A_0(T)$ to indicate that we have kept terms of order T^0, when replacing Q_T by the right-hand side. We apply these expansions to approximate $E(Q_T)$ by the expectation of the right-hand side and write therefore

$$E[Q_T] \overset{1}{=} E[A_0(T)] + T^{-1/2}E[A_1(T)] + T^{-1}E[A_2(T)]$$

or $E[Q_T] \overset{0}{=} E[A_0(T)]$ depending on the circumstances. Thus the expansions of the expectations are formal in the sense that we do not prove that in fact the difference between left and right-hand side of the expressions for the expectations is of smaller order than indicated in the approximation. In a few occasions I have replaced a long calculation of a second moment by noting that some variables are asymptotically independent. This saves a lot of space and I still use the notation $\overset{0}{=}$.

The next section gives the main analytic result and a suggestion for a correction factor. Section 3 contains some examples where the correction can be worked out explicitly and the results are illustrated by simulation. Section 4 concludes. Appendix A contains the details of the proof, which rely on two previous papers by Johansen (2000, 2002), where a correction factor is derived for hypotheses on β.

2. THE MAIN RESULT

In this section the main result is stated in Theorem 1, and the practical implementation of the correction is discussed in Section 2.5. Before formulating the result we need a discussion of the deterministic terms, some notation for functions of the parameters, and some product moment matrices derived from a random walk.

2.1. *The Models and the Deterministic Terms*

Under the null hypothesis $\Pi = \alpha\beta'$, $\Upsilon = \alpha\rho'$ the model $\mathcal{M}_1$ becomes

$$\mathcal{M}_2 : \Delta X_t = \alpha(\beta' X_{t-1} + \rho' t^{n_d}) + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \sum_{i=0}^{n_d-1} \Phi_i t^i + \varepsilon_t, \tag{5}$$

where α and β are $(n \times r)$ and ρ is $(1 \times r)$. The models we get for $n_d = 0, 1$, and perhaps even $n_d = 2$, are used in practice and the notation is chosen to cover such cases. This model corresponds to the trend t^{n_d} being restricted to the cointegration space. We use the notation $D_t = t^{n_d}$, $d_t = (1, \ldots, t^{n_d-1})$, and $\Phi d_t = \sum_{i=0}^{n_d-1} \Phi_i t^i$. We note that $d_{t+1} = Ld_t$, where for $n_d = 3$, say,

$$d_{t+1} = \begin{pmatrix} 1 \\ t+1 \\ (t+1)^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 2 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ t \\ t^2 \end{pmatrix} = Ld_t. \tag{6}$$

Note that the eigenvalues of L are one and that $\operatorname{tr}\{L^m\} = n_d$. Other deterministic terms can be treated in the same fashion but the formulae become more complicated. It turns out that all the results remain valid for the model without deterministic terms, and we shall give the formulae for the correction also for this case. It is not difficult to see that the relation (6) implies that $E(\beta' X_t + \rho' t^{n_d})$ and $E(\Delta X_t)$ are linear in d_t; see Johansen (2002).

2.2. *Parameter Functions*

Under the assumption that the process X_t, given by (5), is $I(1)$, we let Y_t be the process

$$(X_t'\beta, \Delta X_t', \Delta X_{t-1}', \dots, \Delta X_{t-k+2}')' \tag{7}$$

corrected for its mean. The process Y_t is a stationary AR(1) process of dimension $n_y = r + (k-1)n$, given by the equations

$$Y_t = PY_{t-1} + Q\varepsilon_t$$

with

$$P = \begin{pmatrix} I_r + \beta'\alpha & \beta'\Gamma_1 & \cdots & \beta'\Gamma_{k-2} & \beta'\Gamma_{k-1} \\ \alpha & \Gamma_1 & \cdots & \Gamma_{k-2} & \Gamma_{k-1} \\ 0 & I_n & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & I_n & 0 \end{pmatrix}, \quad Q = \begin{pmatrix} \beta' \\ I_n \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{8}$$

We use the identity

$$\alpha(\alpha'\Omega^{-1}\alpha)^{-1}\alpha'\Omega^{-1} + \Omega\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}\alpha_\perp' = I_n, \tag{9}$$

to decompose ε_t into the permanent shocks B_t and transitory shocks U_t:

$$B_t = (\alpha_\perp'\Omega\alpha_\perp)^{-\frac{1}{2}}\alpha_\perp'\varepsilon_t, \quad U_t = (\alpha'\Omega^{-1}\alpha)^{-\frac{1}{2}}\alpha'\Omega^{-1}\varepsilon_t \tag{10}$$

of dimensions $n_b = n - r$ and $n_u = r$, so that $(B_t', U_t')'$ are i.i.d. $N_n(0, I_n)$. We find the representation

$$Y_t = \sum_{m=0}^{\infty} P^m Q\varepsilon_{t-m} = \sum_{m=0}^{\infty}(\theta_m U_{t-m} + \psi_m B_{t-m}) = Y_{\theta t} + Y_{\psi t},$$

with

$$\theta_m = P^m Q\alpha(\alpha'\Omega^{-1}\alpha)^{-\frac{1}{2}}, \quad \psi_m = P^m Q\Omega\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-\frac{1}{2}}. \tag{11}$$

We define the variances and covariances

$$\begin{aligned}
(12)\qquad \Sigma_\theta &= \mathrm{var}(Y_{\theta t}) = \sum_{m=0}^{\infty} \theta_m \theta_m', \\
\Sigma_\psi &= \mathrm{var}(Y_{\psi t}) = \sum_{m=0}^{\infty} \psi_m \psi_m', \\
\Sigma &= \mathrm{var}(Y_t) = \sum_{m=0}^{\infty} (\theta_m \theta_m' + \psi_m \psi_m') = \Sigma_\theta + \Sigma_\psi, \\
\gamma_\theta(h) &= \mathrm{cov}(Y_{\theta t}, Y_{\theta, t+h}) = \Sigma_\theta P'^h, \\
\gamma_\psi(h) &= \mathrm{cov}(Y_{\psi t}, Y_{\psi, t+h}) = \Sigma_\psi P'^h, \\
\gamma(h) &= \mathrm{cov}(Y_t, Y_{t+h}) = \Sigma P'^h = \gamma_\psi(h) + \gamma_\theta(h).
\end{aligned}$$

The variance Σ can be found from the linear equations $\Sigma = P\Sigma P' + Q\Omega Q'$, with solution $\mathrm{vec}(\Sigma) = (I_{n_y^2} - P \otimes P)^{-1}\mathrm{vec}(Q\Omega Q')$. For numerical purposes it is an advantage to diagonalize P and let $P = KRK^{-1}$, where $R = \mathrm{diag}(\rho_1, \dots, \rho_{n_y})$. Then

$$K^{-1}\Sigma K'^{-1} = RK^{-1}\Sigma K'^{-1}R + K^{-1}Q\Omega Q'K'^{-1},$$

and hence $\Sigma = K\{(K^{-1}Q\Omega Q'K'^{-1})_{ij}/(1-\rho_i\rho_j)\}K'$. The long-run coefficients

$$\begin{aligned}
(13)\qquad \theta &= \sum_{m=0}^{\infty} \theta_m = (I_{n_y} - P)^{-1} Q\alpha(\alpha'\Omega^{-1}\alpha)^{-\frac{1}{2}}, \\
\psi &= \sum_{m=0}^{\infty} \psi_m = (I_{n_y} - P)^{-1} Q\Omega\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-\frac{1}{2}},
\end{aligned}$$

give the long-run variances $\theta\theta'$ and $\psi\psi'$ of $Y_{\theta t}$ and $Y_{\psi t}$ respectively. We normalize these by Σ, the variance of Y_t, and define the matrices

$$\begin{aligned}
(14)\qquad V_\theta &= \theta\theta'\Sigma^{-1} = (I_{n_y} - P)^{-1} Q\alpha(\alpha'\Omega^{-1}\alpha)^{-1}\alpha'Q'(I_{n_y} - P')^{-1}\Sigma^{-1}, \\
V_\psi &= \psi\psi'\Sigma^{-1} = (I_{n_y} - P)^{-1} Q\Omega\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}\alpha_\perp'\Omega Q'(I_{n_y} - P')^{-1}\Sigma^{-1}.
\end{aligned}$$

Finally we need matrix

$$(15)\qquad V = \psi'\Sigma^{-1}\sum_{m=0}^{\infty}\psi_m \mathrm{tr}\{\Sigma^{-1}\gamma(m+1)\} + \psi'\Sigma^{-1}\sum_{m=0}^{\infty}\gamma(m+1)'\Sigma^{-1}\psi_m.$$

The trace of V can be calculated as follows:

$$\begin{aligned}
&\mathrm{tr}\left\{\psi'\Sigma^{-1}\sum_{m=0}^{\infty}\psi_m\right\}\mathrm{tr}\{\Sigma^{-1}\gamma(m+1)\}\\
&\quad=\sum_{m=0}^{\infty}\mathrm{tr}\{Q\Omega\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}\alpha_\perp'\Omega Q'(I_{n_y}-P')^{-1}\Sigma^{-1}P^m\}\mathrm{tr}\{P^{m+1}\}\\
&\quad=\sum_{m=0}^{\infty}\mathrm{tr}\{(I_{n_y}-P)V_\psi P^m\}\mathrm{tr}\{P^{m+1}\}\\
&\quad=\sum_{m=0}^{\infty}\mathrm{tr}\{[(I_{n_y}-P)V_\psi\otimes P][P^m\otimes P^m]\}\\
&\quad=\mathrm{tr}\{[(I_{n_y}-P)V_\psi\otimes P][I_{n_y^2}-P\otimes P]^{-1}\},
\end{aligned}$$

and

$$\begin{aligned}
&\mathrm{tr}\left\{\psi'\Sigma^{-1}\sum_{m=0}^{\infty}\gamma(m+1)'\Sigma^{-1}\psi_m)\right\}\\
&\quad=\sum_{m=0}^{\infty}\mathrm{tr}\{Q\Omega\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}\alpha_\perp'\Omega Q'(I_{n_y}-P')^{-1}\Sigma^{-1}P^{2m+1}\}\\
&\quad=\mathrm{tr}\{(I_{n_y}-P)V_\psi P(I_{n_y}-P^2)^{-1}\}=\mathrm{tr}\{V_\psi P(I_{n_y}+P)^{-1}\}.
\end{aligned}$$

Hence we find

$$\mathrm{tr}\{V\}=\mathrm{tr}\{[(I_{n_y}-P)V_\psi\otimes P][I_{n_y^2}-P\otimes P]^{-1}\}+\mathrm{tr}\{V_\psi P(I_{n_y}+P)^{-1}\}. \tag{16}$$

Equivalently we can find an expression in terms of the eigenvectors and eigenvalues of $P=KRK^{-1}$:

$$\mathrm{tr}\{V\}=\sum_{i=1}^{n_y}\sum_{j=1}^{n_y}(K^{-1}V_\psi K)_{ii}\frac{\rho_j(1-\rho_i)}{1-\rho_i\rho_j}+\sum_{i=1}^{n_y}(K^{-1}V_\psi K)_{ii}\frac{\rho_i}{1+\rho_i}. \tag{17}$$

2.3. *Product Moments of Random Walks*

We define the extended process A_t of dimension $n_a=n-r+1$,

$$A_{t-1}=\left(\begin{array}{c|c}\sum_{i=1}^{t-1}B_i & \\ & 1,\ldots,t^{n_d-1}\\ t^{n_d} & \end{array}\right),$$

that is, $\sum_{i=1}^{t-1}B_i$ and t^{n_d} corrected for $1,\ldots,t^{n_d-1}$, and the product moments

$$M_{aa}=\sum_{t=1}^{T}A_{t-1}A_{t-1}',\quad M_{ab}=\sum_{t=1}^{T}A_{t-1}B_t',\quad M_{ab}^{+}=\sum_{t=1}^{T}A_{t-1}B_{t-1}', \tag{18}$$

$$M_{bb}(j)=\sum_{t=1}^{T}B_{t-j}B_t',\ M_{bb}=M_{bb}(0),$$

and use the notation $M_{bb.d} = M_{bb} - M_{bd} M_{dd}^{-1} M_{db}$. We need a number of moments defined from A_{t-1} and B_t:

$$
\begin{aligned}
(19) \qquad & M = M_{ba.d} M_{aa.d}^{-1} M_{ab.d}, \\
& M_{+} = M_{ba.d} M_{aa.d}^{-1} M_{ab.d}^{+}, \\
& M_{++} = M_{ba.d}^{+} M_{aa.d}^{-1} M_{ab.d}^{+}, \\
& M_{-} = M_{bb}(1)(T^{-1} M_{bb.d})^{-1} M_{+}, \\
& K(j) = (M_{bb}(j+1) - M_{bb}(1))(T^{-1} M_{bb.d})^{-1} M_{+} \qquad (j = 1, \ldots).
\end{aligned}
$$

The moments M, M_{+}, and M_{++} converge to functionals of Brownian motion as discussed in Lemma 2 in the Appendix, whereas $K(j)$ and M_{-} have to be normalized by $T^{-1/2}$ to converge.

2.4. *The Correction Factor*

We now formulate the main result about an approximation of the expectation of the log likelihood ratio test for cointegration.

THEOREM 1: *Under the assumption that X_t is an $I(1)$ process with cointegrating rank r given by model $\mathscr{M}_2$, the expectation of the likelihood ratio test of $\Pi = \alpha\beta'$ and $\Upsilon = \alpha\rho'$ in model* (1), *that is, $\mathscr{M}_2$ in $\mathscr{M}_1$, has the expansion*

$$
\begin{aligned}
(20) \qquad & E_\theta[-2 \log LR(\mathscr{M}_2 | \mathscr{M}_1)] \\
& \stackrel{1}{=} -TE[\log |I_{n_b} - M_{bb.d}^{-1} M_{ba} M_{aa}^{-1} M_{ab}|] \\
& \quad + T^{-1}\big(\mathrm{tr}\{E[M]\}\mathrm{tr}\{\psi' \Sigma^{-1} \psi\} \\
& \qquad + \mathrm{tr}\{E[M_{+}]\}\mathrm{tr}\{I_{n_y} - \theta\theta' \Sigma^{-1} - \psi\psi' \Sigma^{-1}\}\big) \\
& \quad + 2T^{-1}(\mathrm{tr}\{E[M_{+}]V\} + n_d \mathrm{tr}\{E[M_{+}]\psi' \Sigma^{-1} \psi\}) \\
& \quad + T^{-1}\mathrm{tr}\{E[M_{+}' M_{+} - 2M_{-} + (n-r)M_{++}]\psi' \Sigma^{-1} \psi\} \\
& \quad - 2T^{-1}\mathrm{tr}\left\{\psi' \Sigma^{-1} \sum_{j=0}^{\infty} \psi_j E[K(j)]\right\}.
\end{aligned}
$$

Here the moments M, M_{+}, M_{++}, M_{-}, and $K(j)$ are defined in (18) *and* (19) *and the coefficients Σ, θ, ψ, V are found in* (12), (13), *and* (15).

The proof will be given in Appendix A based upon an expansion of the likelihood ratio statistic. We next derive the correction factor. Let $n_b = n - r$ be the dimension of B_t or the number of common trends, and define the functions (see (4))

$$f(T,n_b,n_d) = -TE[\log|I_{n_b} - M_{bb.d}^{-1}M_{ba}M_{aa}^{-1}M_{ab}|] \to f(n_b,n_d), \tag{21}$$

$$a(T,n_b,n_d) = f(T,n_b,n_d)/f(n_b,n_d),$$

$$g(T,n_b,n_d) = n_b \text{tr}\{E[M_+]\}/\text{tr}\{E[M]\} \to g(n_b,n_d),$$

$$h(T,n_b,n_d) = \text{tr}\{E[M_+'M_+ + n_b M_{++} - 2M_-]\}/(n_b\text{tr}\{E[M]\}) \to h(n_b,n_d),$$

$$k(T,n_b,n_d,j) = \text{tr}\{E[K(j)]\}/(n_b\text{tr}\{E[M]\}) \to k(n_b,n_d,j).$$

A consequence of Theorem 1 is the following corollary, which implements the approximation to a correction factor for the likelihood ratio test.

COROLLARY 1: *Under the assumption that X_t is an $I(1)$ process with cointegrating rank r given by model $\mathcal{M}_2$, the correction factor for the test of $\mathcal{M}_2$ in $\mathcal{M}_1$ (see* (4)*), that is, the test for cointegrating rank r in the vector autoregressive model* (1), *is given by $a(T,n_b,n_d)(1+T^{-1}b(\theta))$, where $n_b = n - r$ and*

$$b(\theta) = c_1(1+h(n_b,n_d)) + (n_b c_2 + 2(c_3 + n_d c_1))\frac{g(n_b,n_d)}{n_b^2} - 2\text{tr}\left\{\psi'\Sigma^{-1}\sum_{j=0}^{\infty}\psi_j\right\}k(n_b,n_d,j). \tag{22}$$

Here Σ, ψ, and ψ_j are defined in (11), (12), *and* (13), *and*

$$c_1 = \text{tr}\{V_\psi\} = \text{tr}\{(I_{n_y} - P)^{-1}Q\Omega\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}\alpha_\perp'\Omega Q'(I_{n_y} - P')^{-1}\Sigma^{-1}\},$$

$$c_2 = \text{tr}\{I_{n_y} - V_\theta - V_\psi\} = \text{tr}\{I_{n_y} - (I_{n_y} - P)^{-1}Q\Omega Q'(I_{n_y} - P')^{-1}\Sigma^{-1}\},$$

$$c_3 = \text{tr}\{V\} = \text{tr}\{[(I_{n_y} - P)V_\psi \otimes P][I_{n_y^2} - P\otimes P]^{-1}\} + \text{tr}\{V_\psi P(I_{n_y} + P)^{-1}\}.$$

The functions $a(T,n_b,n_d)$, $g(n_b,n_d)$, $k(n_b,n_d,j)$, and $h(n_b,n_d)$ are defined by (19) *and* (21), *and the matrices V_ψ, V_θ, and V are given in* (14) *and* (15). *The formulae remain valid for the model without deterministic terms, where $d_t = D_t = 0$, with an obvious change of the moments.*

PROOF: Let the terms in (20) be $L_0(1+T^{-1}(L_1+L_2+L_3+L_4+L_5+L_6))$. The first term is $L_0 = f(T,n_b,n_d)$; see (3). Because $T^{-1}M_{bb.d} \overset{0}{=} I_{n_b}$ and $M_{ba}M_{aa}^{-1}M_{ab} = M \in O_P(1)$, we have

$$L_0 = -TE[\log|I_{n_b} - T^{-1}(T^{-1}M_{bb.d})^{-1}M_{ba}M_{aa}^{-1}M_{ab}|] \overset{0}{=} \text{tr}\{E[(T^{-1}M_{bb.d})^{-1}M]\} \overset{0}{=} \text{tr}\{E[M]\}.$$

Hence we find

$$L_1 = \mathrm{tr}\{E[M]\}\mathrm{tr}\{\psi'\Sigma^{-1}\psi\}/f(T, n_b, n_d) \overset{0}{=} c_1,$$
$$L_2 = \mathrm{tr}\{E[M_+]\}\mathrm{tr}\left\{I_{n_y} - \theta\theta'\Sigma^{-1} - \psi\psi'\Sigma^{-1}\right\}/f(T, n_b, n_d)$$
$$\overset{0}{=} c_2 g(n_b, n_d)/n_b.$$

It is shown in Lemma 2 in the Appendix, that the matrices $M, M_+, M_{++}, M_-, K(j)$ have expectations that are proportional to the identity matrix, so that for instance $\mathrm{tr}\{E[M_+]V\} = n_b^{-1}\mathrm{tr}\{E[M_+]\}\mathrm{tr}\{V\}$. It follows that

$$L_3 = 2(\mathrm{tr}\{E[M_+]V\}/f(T, n_b, n_d) \overset{0}{=} 2c_3 g(n_b, n_d)/n_b^2,$$
$$L_4 = 2n_d \mathrm{tr}\{E[M_+]\psi'\Sigma^{-1}\psi\}/f(T, n_b, n_d) \overset{0}{=} 2n_d c_1 g(n_b, n_d)/n_b^2,$$
$$L_5 = \mathrm{tr}\{E[M_+'M_+ - 2M_- + (n-r)M_{++}]\psi'\Sigma^{-1}\psi\}/f(T, n_b, n_d)$$
$$\overset{0}{=} c_1 h(n_b, n_d),$$
$$L_6 = -2\mathrm{tr}\left\{\psi'\Sigma^{-1}\sum_{j=0}^{\infty}\psi_j E[K(j)]\right\}\Big/ f(T, n_b, n_d)$$
$$\overset{0}{=} -2\mathrm{tr}\left\{\psi'\Sigma^{-1}\sum_{j=0}^{\infty}\psi_j\right\} k(n_b, n_d, j),$$

which is the result given in Corollary 1. *Q.E.D.*

The matrix appearing in c_2 is $V_\theta + V_\psi = (\theta\theta' + \psi\psi')\Sigma^{-1}$ which measures the "ratio" of the long-run variance to the short-run variance of the process Y_t. We do not have an interpretation of the coefficient c_3. Note that the parameters c_1, c_2, and c_3 do not depend on Φ or n_d, but on the remaining parameters $(\alpha, \beta, \Gamma_1, \ldots, \Gamma_{k-1}, \Omega)$ and hence the dimensions n and r and lag length k.

2.5. *Implementation of the Correction Factor*

In order to implement the correction (22) in practice we need to calculate the coefficients $f(T, n_b, n_d), a(T, n_b, n_d), g(T, n_b, n_d), h(T, n_b, n_d)$, and $k(T, n_b, n_d, j)$. These are complicated functions of a random walk, and we choose to tabulate them by simulation for the most commonly used values of $n_d = 0, 1, 2$, as well as for the situation with no deterministic terms, denoted by $n_d = *$.

We use 500,000 simulations and fit a polynomial in T^{-1} to $f(T, n_b, n_d)$ to determine the limit $f(n_b, n_d)$. Then $a(T, n_b, n_d) = f(T, n_b, n_d)/f(n_b, n_d)$ is described by a polynomial in n_b/T and $1/T$; see Table I. This approximation is correct to two decimals. A similar methodology is used to find $g(n_b, n_d)$ and $h(n_b, n_d)$ (see Table II), and to describe them as a polynomial in $1/n_b$. Finally it is seen by the same methods that $k(T, n_b, n_d, j) \to 0, T \to \infty$, so that $k(n_b, n_d, j) = 0$ is a good approximation.

TABLE I

AN APPROXIMATION OF $a(T, n_b, n_d)$

n_d	$a(T, n_b, n_d)$[a]			
	$a_1(n_d)$	$a_2(n_d)$	$a_3(n_d)$	$b(n_d)$
$*$	0.561	−0.016	2.690	−0.569
0	0.494	0.826	0.829	−0.200
1	0.541	0.625	1.077	−1.518
2	0.570	−0.024	2.645	−2.609

[a]The approximation of $a(T, n_b, n_d)$ is of the form $1 + a_1(n_d)n_b/T + a_2(n_d)(n_b/T)^2 + a_3(n_d)(n_b/T)^3 + b_1(n_d)T_d^{-1}$, for $n_d = *, 0, 1, 2$, where $*$ denotes the model without deterministic terms.

Details on the simulations can be found in Johansen, Hansen, and Fachin (2002), where also a RATS program is given for calculating the correction factor for given values of (θ, T, n_b, n_d). We therefore propose to use the correction factor $a(T, n_b, n_d)(1 + b(\hat{\theta})/T)$, where

$$b(\theta) = c_1(1 + h(n_b, n_d)) + (n_b c_2 + 2(c_3 + n_d c_1))\frac{g(n_b, n_d)}{n_b^2}, \tag{23}$$

where the approximations to the coefficients $a(T, n_b, n_d), g(n_b, n_d)$, and $h(n_b, n_d)$ are given in Tables I and II, and the coefficients $c_i = c_i(\theta)$ are defined in Corollary 1.

The next section contains some simulations to investigate if the correction factor can improve the asymptotic tables for the likelihood ratio test to get more accurate information on the size of the test. Here we briefly discuss the power.

It is obvious that when correcting the size of a nominal 5% test from the actual value around 25%, say, to around 5% by the correcting factor, the power is decreased correspondingly. In order to investigate if the correction factor also improves the finite sample approximation under the alternative one would have to calculate the expectation under the alternative, and the reformulations and expansions found in this paper may have to be modified. We know, however,

TABLE II

AN APPROXIMATION OF $h(n_b, n_d)$ AND $g(n_b, n_d)$

n_d	$h(n_b, n_d)$[a]			$g(n_b, n_d)$[a]			
	$h_1(n_d)$	$h_2(n_d)$	$h_3(n_d)$	$g_0(n_d)$	$g_1(n_d)$	$g_2(n_d)$	$g_3(n_d)$
$*$	0.000	0.000	0.000	−0.506	0.020	0.070	−0.144
0	0.000	0.197	0.036	−0.496	0.166	0.079	−0.076
1	0.000	3.218	−1.401	−1.499	1.663	−1.091	0.304
2	0.200	7.699	−4.214	−2.459	4.528	−4.848	2.049

[a]The approximation of $h(n_b, n_d)$ is of the form $h_1(n_d)n_b^{-1} + h_2(n_d)n_b^{-2} + h_3(n_d)n_b^{-3}$ and the approximation of $g(n_b, n_d)$ has the form $g_0(n_d) + g_1(n_d)n_b^{-1} + g_2(n_d)n_b^{-2} + g_3(n_d)n_b^{-3}$ for $n_d = *, 0, 1, 2$, where $*$ denotes the model without deterministic terms.

that the likelihood ratio test is consistent in the sense that if the data generating process has rank $r+r_1$, then $-2\log LR$ is of the order of T and hence the power tends to one. The asymptotic power against local alternatives has been studied in Johansen (1991). The corrected test will be consistent with the same local power properties if $b(\hat{\theta})$ is bounded. The coefficient $b(\theta)$ is a function of the parameters as given in Theorem 1, and it is seen that as long as $(I_{n_y}-P)$ is a full rank matrix, or equivalently P does not have a unit root, then the coefficients are well behaved. In order to analyze this, assume that the DGP under the alternative has the form

$$\Delta X_t = \alpha(\beta' X_{t-1} + \rho' t^{n_d}) + \alpha_1(\beta_1' X_{t-1} + \rho_1' t^{n_d}) + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \sum_{i=1}^{n_d-1} \Phi_i t^i + \varepsilon_t, \tag{24}$$

where β is $(n\times r)$ and β_1 is $(n\times r_1)$. Under the local alternative we assume that α_1 is $O(T^{-1})$. Let $P=P(r)$ be defined by (8) on the basis of only α and β, whereas $P(r+r_1)$ is defined similarly on the basis of (β,β_1) and (α,α_1). The assumption that X_t is $I(1)$ under the alternative means that $P(r+r_1)$ has eigenvalues less than 1 in absolute value, and the assumption that $X(t)$ is $I(1)$ under the null hypothesis means that $P(r)$ has eigenvalues less than 1 in absolute value. We want to estimate the parameters of the null hypothesis so that they are consistent even under the local alternative. We therefore estimate them from the unrestricted VAR by reduced rank regression. By the methods of Johansen (1996, Chapter 14) it can be proved that these estimators are consistent even under the local alternative.

3. SOME SPECIAL CASES

In this section we illustrate the results in some special cases, where the coefficients can be worked out explicitly, and which are therefore convenient for gaining some intuition for the result.

3.1. *The Test for No Cointegration in the Model with Two Lags*

We consider the test of $\Pi=0$ and $\Upsilon=0$ in the model (1) with $k=2$ and $n_d=1$:

$$\Delta X_t = \Pi X_{t-1} + \Gamma_1 \Delta X_{t-1} + \Upsilon t + \Phi + \varepsilon_t. \tag{25}$$

Under the null hypothesis there is no cointegration, but parameters Γ_1, Φ, and Ω. The distribution of the test statistic does not depend on Φ and Ω, so in the DGP we can take $\Phi=0$ and $\Omega=I_n$. We can find a simple expression for the correction factor, if we consider the DGP given by $\Gamma_1=\xi I_n$, $-1<\xi<1$, and hence see the effect of the short term dynamics. In this case we have $\alpha=\beta=0$, $\alpha_\perp=\beta_\perp=I_n$,

TABLE III

TEST FOR $\Pi = 0$ AND $\Upsilon = 0$ IN MODEL (25)[a]

n	n_d	$T\backslash\xi$	0	0.3	0.5	0.6	0.7	0.9
5	1	50	$\frac{37.6}{11.4}$(1.14)	$\frac{55.7}{10.5}$(1.22)	$\frac{73.7}{7.2}$(1.34)	$\frac{82.7}{4.2}$(1.43)	$\frac{92.0}{1.2}$(1.60)	$\frac{98.2}{0.0}$(2.94)
5	1	100	$\frac{16.8}{8.0}$(1.07)	$\frac{24.6}{7.5}$(1.11)	$\frac{35.2}{7.2}$(1.16)	$\frac{44.2}{6.0}$(1.21)	$\frac{59.0}{4.3}$(1.30)	$\frac{94.1}{0.0}$(1.95)

[a]For the simulations we assume $\Pi = 0$, $\Upsilon = 0$, $\Gamma_1 = \xi I_n$, and $\Omega = I_n$. Each entry shows the simulated rejection probability of a nominal 5% test using asymptotic critical values, over the rejection probability for the corrected test with the correction factor in parenthesis. The number of simulations is 10,000. The correction is calculated from (26).

and that $Y_t = \Delta X_t - E(\Delta X_t)$ is autoregressive with coefficient $P = \xi I_n$, $Q = I_n$, and $Y_t = \sum_{m=0}^{\infty} \xi^m \varepsilon_{t-m}$. This gives

$$\Sigma = \operatorname{var}(\Delta X_t) = \sum_{m=0}^{\infty} \xi^{2m} \Omega = \frac{1}{1-\xi^2}\Omega, \quad V_\psi = \frac{1+\xi}{1-\xi} I_n, \quad V_\theta = 0,$$

so that $c_1 = n(1+\xi)/(1-\xi)$, $c_2 = -2n\xi/(1-\xi)$, $c_3 = n(n+1)\xi/(1-\xi)$. From (23) we get the correction factor with $k = 2$, $r = 0$, $n_b = n$, $n_d = 1$:

$$(26) \qquad a(T, n, 1)\left(1 + \frac{1}{T(1-\xi)}\left[n(1+\xi)(1+h(n,1)) + 2\frac{1}{n}(2\xi+1)g(n,1)\right]\right).$$

Some simulations were performed and are given in Table III to illustrate the usefulness of formula (26). Note that as ξ tends to 1, where X_t becomes $I(2)$, the rejection probability of the test increases almost to one, and hence the asymptotic $I(1)$ critical values (Johansen (1996, Table 15.4)) for the trace test are not useful. Thus, it is important to test if the process is $I(2)$ before determining the rank of Π. The correction factor manages to correct the rejection probability to a reasonable level for $\xi \leq 0.6 - 0.7$, say. For $\xi = 0.6$ and $T = 50$ a nominal 5% test, using the asymptotic critical values, has a rejection probability of 83%. The correction brings the value down to around 4%. As the test becomes even more distorted the correction factor overcorrects due to the singularity in the expression $1/(1-\xi)$.

If we define the size of the test of $\Pi = 0$, $\Upsilon = 0$, using a critical region C, as $\max_{\Gamma_1} P_{\Gamma_1}(C)$, then by applying the asymptotic critical values the size of the test is very close to 1, since this is the value we get for ξ close to 1. The corrected test appears to attain the largest rejection probability of 7–9% around $\xi = 0$, but the choice of DGP ($\Gamma_1 = \xi I_n$) does not allow determination of the actual size.

3.2. *The Test for Rank One in the Model with One Lag*

The model with $k = 1$ and $n_d = 1$ is

$$(27) \qquad \Delta X_t = \Pi X_{t-1} + \Upsilon t + \Phi + \varepsilon_t,$$

and we test $\Pi = \alpha\beta'$, $\Upsilon = \alpha\rho$, where α and β are $(n\times 1)$. Under the null hypothesis,

$$Y_t = \beta' X_t - E(\beta' X_t) = \sum_{i=0}^{\infty} (1+\beta'\alpha)^i \beta' \varepsilon_{t-i},$$

so that $\Sigma = \mathrm{var}(\beta' X_t) = \beta'\Omega\beta/(1-(1+\beta'\alpha)^2)$ and $P = 1+\beta'\alpha$, $Q = \beta'$, $n_y = 1$. We define the parameter

$$\kappa = \frac{\beta'\Omega\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}\alpha_\perp'\Omega\beta}{\beta'\Omega\beta} = 1 - \frac{(\beta'\alpha)^2}{\alpha'\Omega^{-1}\alpha\beta'\Omega\beta}$$

(see (9)), and find the coefficients $c_1 = -(2+\beta'\alpha)\kappa/\beta'\alpha$, $c_2 = 2(1+\beta'\alpha)/\beta'\alpha$, $c_3 = -2(1+\beta'\alpha)\kappa/\beta'\alpha$. From (23) we find the correction factor for testing $r=1$ in the model with only one lag:

$$a(T, n-1, 1)\left(1 + \frac{1}{T\beta'\alpha}\left[-(2+\beta'\alpha)\kappa(1+h(n-1,1)) + \{2(1+\beta'\alpha)(n-1) - 2\kappa(4+3\beta'\alpha)\}\frac{g(n-1,1)}{(n-1)^2}\right]\right). \tag{28}$$

In model (27) it is enough to consider DGP's with $\Omega = I_n, \beta = (1, 0, \dots, 0), \alpha = (a_1, a_2, 0, \dots, 0)$ so that $\alpha'\beta = a_1$. The stationary process $\beta' X_t - E(\beta' X_t)$ is autoregressive with parameter $1+\beta'\alpha = 1+a_1$. If we consider a sequence $\alpha_\lambda \to \alpha = 0$, then $a_1 = \beta'\alpha \to 0$, and in the limit $r = 0$, so there is no cointegration. If, however, we let $\alpha_\lambda \to \alpha \neq 0$, but so that $\alpha'\beta = 0$, then again $a_1 = \beta'\alpha \to 0$ but the process is in the limit $I(2)$. Thus the interpretation of the singularity at $a_1 = \alpha'\beta = 0$ depends on the direction it is approached.

We choose to tabulate the rejection probability as a function of (a_1, a_2) for $a_1 = -0.1, \dots, -0.8, a_2 = 0.0, -0.1, \dots, -0.8$. The function is symmetric in a_2, so that Table IV suitably extended covers all possible simulation results for $n = 5, r = 1$. It is seen from Table IV that if we are close to an $I(1)$ model with lower rank $(a_1 \to 0, a_2 = 0)$, then the distribution is shifted to the left, and the rejection probability is less than 5%, and if we are close to an $I(2)$ model $(a_1 \to 0, a_2 \neq 0)$, the distribution is shifted to the right, and the rejection probability is greater than 5%, as was also seen in Section 3.1.

By applying the asymptotic critical values the size of the test is at least 29.9% for $T = 50$, since this is the value we get for the combination $a_1 = -0.1$ and $a_2 = -0.8$. By applying the correction we get a test where the maximal simulated value of the rejection probability is attained for $a_1 = a_2 = -0.8$, and further simulations indicate a size around 9%.

3.3. *The Dickey Fuller Test for Rank Zero in the Model with k Lags*

We consider the test of $\Pi = 0$ and $\Upsilon = 0$ in the statistical model (1). We can evaluate the expectation of the test statistic in the simple case of a DGP, where

TABLE IV

TEST FOR $\Pi = \alpha\beta'$, $\Upsilon = \alpha\rho$ IN MODEL (27)[a]

$T = 50$				
$a_2 \backslash a_1$	−0.1	−0.2	−0.4	−0.8
0.0	$\frac{2.1}{0.4}$ (1.08)	$\frac{2.2}{0.9}$ (1.05)	$\frac{3.8}{2.3}$ (1.03)	$\frac{8.5}{6.9}$ (1.02)
−0.1	$\frac{2.8}{0.0}$ (1.30)	$\frac{2.8}{0.7}$ (1.10)	$\frac{4.1}{2.1}$ (1.04)	$\frac{9.3}{7.4}$ (1.02)
−0.2	$\frac{6.3}{0.0}$ (1.40)	$\frac{4.7}{0.5}$ (1.15)	$\frac{5.3}{2.6}$ (1.05)	$\frac{9.0}{7.0}$ (1.02)
−0.4	$\frac{20.3}{0.0}$ (1.45)	$\frac{12.8}{1.2}$ (1.20)	$\frac{9.4}{3.7}$ (1.07)	$\frac{10.8}{8.1}$ (1.03)
−0.8	$\frac{29.9}{0.2}$ (1.47)	$\frac{23.7}{2.7}$ (1.22)	$\frac{17.5}{6.8}$ (1.10)	$\frac{13.3}{9.2}$ (1.03)
$T = 100$				
−0.1	$\frac{3.1}{0.2}$ (1.15)	$\frac{2.8}{1.3}$ (1.05)	$\frac{6.1}{4.7}$ (1.02)	$\frac{8.5}{7.6}$ (1.01)
−0.2	$\frac{8.9}{0.6}$ (1.20)	$\frac{6.5}{2.3}$ (1.07)	$\frac{6.9}{4.9}$ (1.02)	$\frac{8.4}{7.6}$ (1.01)
−0.4	$\frac{18.3}{1.5}$ (1.22)	$\frac{12.9}{4.3}$ (1.10)	$\frac{9.3}{6.2}$ (1.03)	$\frac{8.4}{7.5}$ (1.01)
−0.8	$\frac{21.7}{1.6}$ (1.23)	$\frac{15.5}{5.1}$ (1.11)	$\frac{11.6}{7.4}$ (1.05)	$\frac{9.6}{8.2}$ (1.02)

[a]For the simulations we assume that $\Pi = \alpha\beta'$, $\Upsilon = \alpha\rho$, where $\rho = 0$, $\alpha = (a_1, a_2, 0, 0, 0)'$, $\beta = (1, 0, 0, 0, 0)'$, $\Phi = 0$, and $\Omega = I_5$. Each entry shows the simulated rejection probability of a nominal 5% test using asymptotic critical values, over the rejection probability for the corrected test with the correction factor in parenthesis. The number of simulations is 10,000. The correction is calculated from (28).

we assume $\Gamma_1 = \cdots = \Gamma_{k-1} = \Pi = 0$ and $\Upsilon = 0$, in which case $\Delta X_t = \Phi d_t + \varepsilon_t$, in order to see the effect of lag length k. The stationary process Y_t is of dimension $n_y = (k-1)n$, and when all $\Gamma_i = 0$, it is given by

$$Y_t' = (\Delta X_t', \ldots, \Delta X_{t-k+2}') - E(\Delta X_t', \ldots, \Delta X_{t-k+2}') = (\varepsilon_t', \ldots, \varepsilon_{t-k+2}').$$

We find

$$\Sigma = \operatorname{var}(Y_t) = \Omega \otimes I_{k-1}, \quad P = I_n \otimes E_{k-1}, \quad Q = I_n \otimes e_1,$$

where e_i is the ith unit vector in R^{k-1} and E_{k-1} is the $(k-1)\times(k-1)$ shift matrix defined by

$$E_{k-1} = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix} = \begin{cases} \sum_{i=1}^{k-1} e_{i+1}e_i', & k \geq 3, \\ 0, & k = 1, 2. \end{cases}$$

Note that $(E_{k-1})^{k-1} = 0$, and that

$$P^i Q = (I_n \otimes E_{k-1}^i)(I_n \otimes e_1) = (I_n \otimes e_{i+1}) \qquad (i = 0, \ldots, k-1),$$

$$\left(I_{n_y} - P\right)^{-1} Q = I_n \otimes (1, \ldots, 1)' = I_n \otimes \iota,$$

say, where $\iota = \sum_{i=1}^{k-1} e_i$. We find for $\alpha_\perp = I_n$ that $V_\theta = 0$ and

$$\begin{aligned} V_\psi &= \left(I_{n_y} - P\right)^{-1} Q\Omega Q' \left(I_{n_y} - P'\right)^{-1} \Sigma^{-1} \\ &= (I_n \otimes \iota)\Omega(I_n \otimes \iota')(\Omega^{-1} \otimes I_{k-1}) = I_n \otimes \iota\iota', \end{aligned}$$

and hence we get $c_1 = \mathrm{tr}\{V_\psi\} = n(k-1) = n_y$, and since $n_y = \mathrm{tr}\{V_\psi\}$, we find $c_2 = \mathrm{tr}\{I_{n_y} - V_\theta - V_\psi\} = \mathrm{tr}\{I_{n_y} - V_\psi\} = n_y - n_y = 0$. Finally to find c_3 we evaluate

$$\begin{aligned}
&\mathrm{tr}\{[(I_{n_y} - P)V_\psi \otimes P][I_{n_y^2} - P\otimes P]^{-1}\}\\
&\quad = \sum_{i=0}^{\infty} \mathrm{tr}\{[(I_{n_y} - P)V_\psi \otimes P][P^i \otimes P^i]\}\\
&\quad = \sum_{i=0}^{\infty} \mathrm{tr}\{(I_{n_y} - P)V_\psi P^i \otimes P^{i+1}\}\\
&\quad = \sum_{i=0}^{\infty} \mathrm{tr}\{(I_{n_y} - P)V_\psi P^i\}\mathrm{tr}\{P^{i+1}\} = 0,
\end{aligned}$$

since $\mathrm{tr}\{P^{i+1}\} = 0$, and hence

$$\begin{aligned}
c_3 &= \mathrm{tr}\{V_\psi P(I_{n_y} + P)^{-1}\} = \sum_{i=0}^{\infty} \mathrm{tr}\{[I_n \otimes \iota\iota'][I_n \otimes E_{k-1}][I_n \otimes (-1)^i E_{k-1}^i]\}\\
&= \sum_{i=0}^{\infty} (-1)^i \mathrm{tr}\{I_n \otimes \iota\iota' E_{k-1}^{i+1}\} = \sum_{i=0}^{\infty} (-1)^i n\iota' E_{k-1}^{i+1}\iota\\
&= n\sum_{i=0}^{k-2} (-1)^i (k-2-i) = n\left[\frac{k-1}{2}\right],
\end{aligned}$$

that is, n times the integer part of $(k-1)/2$. Thus $c_1 = (k-1)n$, $c_2 = 0$, $c_3 = n[(k-1)/2]$. From (23) we get the result

$$a(T, n, n_d)\left(1 + \frac{1}{T}\left[(k-1)n(1 + h(n, n_d)) + \left\{\left[\frac{k-1}{2}\right] + n_d(k-1)\right\}\frac{2g(n, n_d)}{n}\right]\right). \tag{29}$$

This coincides with (26) for $k = 2$, $\xi = 0$, and $n_d = 1$. For $k = 1$ we just get the correction factor $a(T, n, n_d)$, corresponding to the normalization on the Dickey Fuller test in the model with one lag.

The simulations in Table V show that as long as the number of parameters per observation, kn/T, is less than 0.2, the formula gives a reasonable result. Thus for instance for $n = 5$, $k = 2$, $T = 50$ the rejection probability of a nominal 5% test using asymptotic critical values is in fact close to 37%. The correction gives a test with rejection probability around 11%, which is much better than the direct use of the asymptotic tables.

Note that if we set $h = g = 0$, the second factor becomes $1 + (k-1)n/T$ which corresponds to multiplying the likelihood ratio test by $(T - (k-1)n)/T$, which is the correction found by Hansen and Rahbek (2002) based on an argument involving a profile likelihood, whereas Ahn and Reinsel (1990) and Reimers (1992) suggested use of $(T - kn)/T$. Both of these "degrees of freedom" corrections capture part of the dependence on lag length, but not the dependence on the parameters.

TABLE V

TEST FOR $\Pi = 0$, $\Upsilon = 0$ IN MODEL (1)[a]

$k\backslash T$	50	100	500
1	$\frac{13.5}{9.2}$(1.03)	$\frac{9.4}{8.2}$(1.01)	$\frac{7.3}{6.9}$(1.01)
2	$\frac{37.1}{11.0}$(1.14)	$\frac{16.9}{7.9}$(1.07)	$\frac{8.8}{7.2}$(1.01)
3	$\frac{69.1}{18.1}$(1.23)	$\frac{27.8}{9.4}$(1.11)	$\frac{9.8}{7.5}$(1.02)
4	$\frac{92.7}{31.7}$(1.34)	$\frac{43.8}{11.7}$(1.16)	$\frac{10.3}{6.8}$(1.03)

[a] For the simulation we assume that $n = n_b = 5$, $n_d = 1$, and $\Pi = 0$, $\Upsilon = 0$, $\Gamma_i = 0$, $i = 1, \ldots, k$, $\Omega = I_n$. Each entry shows the simulated rejection probability of a nominal 5% test using asymptotic critical values, over the rejection probability for the corrected test with the correction factor in parenthesis. The number of simulations is 10,000. The correction is calculated using (29).

3.4. *A Real Life Example*

We illustrate the methods by a data set taken from Johansen (1996). We consider the Danish data set consisting of the four variables m_t (log real M2), y_t (log real income), i_t^b (bond rate), and finally i_t^d (deposit rate) observed quarterly from 1974:1 to 1987:3. We fitted a model with two lags, restricted constant term, and seasonal dummies.

We decided in the book to take $r = 1$, to illustrate the methods, even though the trace statistic (49.14) was below the 95% critical value 53.42 in the asymptotic distribution. We here investigate by simulation the rejection probability of the test that uses the asymptotic distribution, and the effect of applying the correction factor in order to mimic an actual application.

In each simulation we use the estimated values of the parameters $(\hat{\alpha}, \hat{\beta}, \widehat{\Gamma}_1, \hat{\rho}, \widehat{\Omega})$ from the Danish data to define the data generating process. We simulate 10,000 time series with 53 observations, which was the number of observations in the example. The processes are started at the actual initial values and, to simplify, the seasonal dummies have been left out. In each simulation we estimate the parameters and the correction factor.

We first let $r = 0$, and hence use only the parameters $(\widehat{\Gamma}_1, \widehat{\Omega})$ from the data generating process. We next assume that $r = 1$ and simulate the data using the estimated adjustment and cointegration vector $\hat{\alpha}$ and $\hat{\beta}$ and $(\widehat{\Gamma}_1, \widehat{\Omega})$. We compare in Table VI the simulated 95% quantiles with the asymptotic ones. The simulated quantiles are larger than the asymptotic ones and this is what is captured by the correction factor. There seems to be very little statistical evidence of cointegration in the Danish data.

We also see that the direct use of the asymptotic tables gives for the test of $r = 0$, a 19% rejection probability instead of the nominal 5%. The rejection probability of the corrected test is 6%. For $r = 1$ we get 17% instead of 5%, but the correction factor brings the rejection probability down to 4%.

When we simulate the DGP with $r = 1$, and use the corrected test statistic for the hypothesis that $r = 0$, we find the power of the test. Using the 95% critical

TABLE VI

THE DANISH DATA[a]

r	$n-r$	95%(asym)	95%(sim)	
0	4	53.42	61.85	$\frac{19.2}{6.4}(1.14)$
1	3	34.80	40.79	$\frac{17.5}{4.2}(1.21)$

[a]The columns give the rank and common trends tested, the asymptotic 95% quantiles and the simulated ones. Next follows the simulated rejection probability of a nominal 5% test using asymptotic critical values, over the rejection probability for the corrected test with the correction factor in parenthesis. As DGP we use the estimated values from the Danish data.

values from the asymptotic distribution we find the power 90.8%, and using the correction factor we get 76.9%. Thus the power function is shifted down by the correction factor. We get the same value whether we use the estimates under the hypothesis $r = 0$ or the estimates from the unrestricted VAR.

4. CONCLUSION

A detailed analysis of the Taylor's expansion of the trace statistic for cointegrating rank gives an approximation of its expectation. This is used to suggest a correction factor to the trace statistic of the form $a(T, n_b, n_d)(1 + T^{-1}b(\hat{\theta}))$, where θ denotes the parameters under the null hypothesis, and $a(T, n_b, n_d)$ is the correction factor needed for the test of no cointegration in the model with one lag and dimension $n_b = n - r$. A numerical approximation to $a(T, n_b, n_d)$ is found by simulation and a computable formula for $b(\theta)$ is given.

A general conclusion from the simulation experiments is that as θ approaches a boundary point, where the process is almost $I(2)$, the rejection probability of a nominal 5% test grows in some cases to almost one, meaning that a nominal 5% test can have very large size. The correction factor also grows and manages to capture part of the size distortion.

Throughout, however, the corrected rejection probability is closer to the nominal value, so there seems to be a large area of the parameter space where the correction appears to be a useful supplement to the tool box for the analysis of cointegrated systems.

Department of Statistics and Operations Research, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen Ø, Denmark; sjo@stat.ku.dk; www.math.ku.dk

Manuscript received July, 2000; final revision received December, 2001.

APPENDIX A

This Appendix contains the proof of Theorem 1. We first show that by introducing a model with a simple hypothesis on the cointegrating space, we can exploit previous results (Johansen (2000, 2002)) and simplify the derivations. Then we introduce a convenient reparameterization of model $\mathscr{M}_1$ and use it to derive an expansion of the test statistic in Theorem 2. Section A.4 contains some asymptotic results on the moment matrices and the detailed evaluation of the terms in the expansion is left to Section A.5.

A.1. *A Simple Hypothesis on the Cointegrating Space*

We define model $\mathcal{M}_3$ by specifying the cointegration space $\mathrm{sp}(\beta^0)$ and $\mathrm{sp}(\rho^0)$, that is $\beta = \beta^0\tau$, $\rho = \rho^0\tau$, $\tau(r \times r)$. In this case the model equations are

$$\mathcal{M}_3\colon\ \Delta X_t = \alpha\tau(\beta^{0\prime}X_{t-1} + \rho^{0\prime}D_t) + \sum_{i=1}^{k-1}\Gamma_i\Delta X_{t-i} + \Phi d_t + \varepsilon_t. \tag{30}$$

Note that the parameter τ can be absorbed in α, and that $\mathcal{M}_3 \subset \mathcal{M}_2 \subset \mathcal{M}_1$, in the sense of inclusion of parameter space.

The test for cointegrating rank is the test of $\mathcal{M}_2$ in $\mathcal{M}_1$, but it is convenient to use the usual trick (see Lawley (1956)) and compare the two models by introducing $\mathcal{M}_3$. Let $L(\mathcal{M}_i)$ denote the maximized likelihood in model $\mathcal{M}_i$; then

$$\frac{L(\mathcal{M}_2)}{L(\mathcal{M}_1)} = \frac{L(\mathcal{M}_3)}{L(\mathcal{M}_i)}\bigg/\frac{L(\mathcal{M}_3)}{L(\mathcal{M}_2)},$$

and hence

$$E_\theta[-2\log LR(\mathcal{M}_2|\mathcal{M}_1)] = E_\theta[-2\log LR(\mathcal{M}_3|\mathcal{M}_1)] + E_\theta[2\log LR(\mathcal{M}_3|\mathcal{M}_2)].$$

The purpose of this expression is that we can use results for $-2\log LR(\mathcal{M}_3|\mathcal{M}_2)$ from Johansen (2000), which deals with the correction factor for a simple hypothesis on β in the cointegration model $\mathcal{M}_2$. Here we find a correction to the joint test of rank and β, $-2\log LR(\mathcal{M}_3|\mathcal{M}_1)$, by deriving an expansion of the expectation and finally we find the required approximation to the expectation of the test for cointegrating rank by subtraction. The reason for introducing the model $\mathcal{M}_3$ is that under the null hypothesis, estimation of (30) is a simple regression that facilitates the calculations.

A.2. *A Reparameterization of* $\mathcal{M}_1$

It is of course easy to derive the test statistic of $\mathcal{M}_3$ in $\mathcal{M}_1$, applying the usual regression formulae based upon equations (1) and (30). We find with an obvious notation

$$-2\log LR(\mathcal{M}_3|\mathcal{M}_1) = -T\log\frac{|M_{\varepsilon\varepsilon\cdot x,\,\Delta x,\,D,\,d}|}{|M_{\varepsilon\varepsilon\cdot\beta^{0\prime}x+\rho^{0\prime}D,\,\Delta x,\,d}|}. \tag{31}$$

We want to calculate the expectation of (31) for a given value of the parameters α^0, β^0, etc. which we call the true value. In order to get more manageable expressions we introduce a new parameterization and new regressors using the true value of the parameters, as described in detail in Johansen (2000, 2002).

We use $\bar\beta_0 = \beta_0(\beta_0'\beta_0)^{-1}$ and $\Gamma^0 = I_n - \sum_{i=1}^{k-1}\Gamma_i^0$, $C^0 = \beta_\perp^0(\alpha_\perp^{0\prime}\Gamma^0\beta_\perp^0)^{-1}\alpha_\perp^{0\prime}$ and define new parameters ψ and $\delta = (\delta_1', \delta_2')'$ as functions of the old (Π, Υ) by

$$\begin{aligned}
\psi' &= \Pi(I_n - C^0\Gamma^0)\bar\beta^0 && (n\times r),\\
\delta_1' &= \Pi\beta_\perp^0 && (n\times(n-r)),\\
\delta_2' &= \Upsilon - \Pi(I_n - C^0\Gamma^0)\bar\beta^0\rho^{0\prime} && (n\times 1).
\end{aligned}$$

The old parameters are given in terms of the new by

$$\Pi = \delta_1'\left(\alpha_\perp^{0\prime}\Gamma^0\beta_\perp^0\right)^{-1}\alpha_\perp^{0\prime}\Gamma^0 + \psi'\beta^{0\prime}, \quad \Upsilon = \delta_2' + \psi'\rho^{0\prime}.$$

The hypothesis $(\Pi, \Upsilon) = (\alpha\tau'\beta^{0\prime}, \alpha\tau'\rho^{0\prime})$ is then expressed as $\delta = 0$.

Model equation (1) with the new parameters is

$$\Delta X_t = \psi'\left(\beta^{0\prime} X_{t-1} + \rho^{0\prime} D_t\right) + \delta_1'\left(\alpha_\perp^{0\prime}\Gamma^0\beta_\perp^0\right)^{-1}\alpha_\perp^{0\prime}\Gamma^0 X_{t-1} + \delta_2' D_t + \sum_{i=1}^{k-1}\Gamma_i \Delta X_{t-i} + \Phi d_t + \varepsilon_t. \tag{32}$$

Under $\mathscr{M}_3$, it holds that $E[\beta^{0\prime}X_{t-1}+\rho^{0\prime}D_t]$ and $E[\Delta X_t]$ are linear in d_t; see Johansen (2002). We therefore introduce the stationary regressors

$$V_{t-1} = \beta^{0\prime}X_{t-1} - E_0(\beta^{0\prime}X_{t-1}),$$
$$Z_{t-1} = (\Delta X_{t-1}' - E_0(\Delta X_{t-1}'), \dots, \Delta X_{t-k+1}' - E_0(\Delta X_{t-k+1}'))',$$

by modifying Φ suitably. Note that $Y_t = (V_t', Z_t')'$; see (7).

We replace the regressors $\left(\alpha_\perp^{0\prime}\Gamma^0\beta_\perp^0\right)^{-1}\alpha_\perp^{0\prime}\Gamma^0 X_{t-1}$ and D_t by

$$A_{t-1} = \left(\begin{array}{c|c} \left(\alpha_\perp^{0\prime}\Omega^0\alpha_\perp^0\right)^{-1/2}\alpha_\perp^{0\prime}\sum_{i=1}^{t-1}\varepsilon_i & d_t \\ D_t & \end{array}\right) = \left(\begin{array}{c|c} \sum_{i=1}^{t-1} B_i & d_t \\ D_t & \end{array}\right), \tag{33}$$

and model equation (32) in the new variables, with suitably redefined parameters Ψ and Φ, becomes

$$\underset{(n)}{\Delta X_t} = \psi' \underset{(r)}{V_{t-1}} + \delta' \underset{(n_a)}{A_{t-1}} + \Psi \underset{((k-1)n)}{Z_{t-1}} + \Phi \underset{(n_d)}{d_t} + \underset{(n)}{\varepsilon_t}, \tag{34}$$

where the dimensions are indicated below each variable. The test for $\mathscr{M}_3$ in $\mathscr{M}_1$ is the test for $\delta = 0$ in (34). The estimators for the parameters ψ, δ, Ψ, Φ, and Ω are found by regression of ΔX_t on $(V_{t-1}, A_{t-1}, Z_{t-1}, d_t)$, and under the hypothesis $\delta = 0$ the parameters can be found by regression of ΔX_t on (V_{t-1}, Z_{t-1}, d_t).

Similarly the model $\mathscr{M}_2$ can be reparameterized as

$$\underset{(n)}{\Delta X_t} = \alpha \underset{(r)}{V_{t-1}} + \alpha\delta' \underset{(n_a)}{A_{t-1}} + \Psi \underset{((k-1)n)}{Z_{t-1}} + \Phi \underset{(n_d)}{d_t} + \underset{(n)}{\varepsilon_t} \tag{35}$$

(see equation (14) in Johansen (2000)) and the test for $\mathscr{M}_2$ in $\mathscr{M}_1$ is the test for $\delta = 0$ in (35).

A.3. *The Likelihood Ratio Test and its Expansion*

We define the product moment matrices $M_{\bullet\bullet}$ for the variables $\Delta X_t, \varepsilon_t$, and d_t at time t but V_{t-1}, A_{t-1}, and Z_{t-1} lagged one period. Thus for instance

$$\sum_{t=1}^{T}\begin{pmatrix}\Delta X_t \\ V_{t-1} \\ \varepsilon_t\end{pmatrix}\begin{pmatrix}\Delta X_t \\ V_{t-1} \\ \varepsilon_t\end{pmatrix}' = \begin{pmatrix} M_{00} & M_{0v} & M_{0\varepsilon} \\ M_{v0} & M_{vv} & M_{v\varepsilon} \\ M_{\varepsilon 0} & M_{\varepsilon v} & M_{\varepsilon\varepsilon}\end{pmatrix}.$$

We use the notation for any three process X_t, U_t, and V_t, say,

$$(U_t|X_t) = U_t - M_{ux}M_{xx}^{-1}X_t, \quad M_{uv\cdot x} = \sum_{t=1}^{T}(U_t|X_t)(V_t|X_t)' = M_{uv} - M_{ux}M_{xx}^{-1}M_{xv},$$

and in particular we use a notation for the moment matrices corrected for the lagged differences Z_{t-1} and d_t, since many results look a bit simpler this way, and some results can be taken from Johansen (2000):

$$S_{uv} = M_{uv\cdot z,d} = M_{uv} - M_{ud}M_{dd}^{-1}M_{dv} - M_{uz\cdot d}M_{zz\cdot d}^{-1}M_{zv\cdot d}. \tag{36}$$

These moment matrices appear naturally when the likelihood function is concentrated with respect to Ψ and Φ.

The likelihood ratio test of $\mathcal{M}_3$ in $\mathcal{M}_1$ is the test that $\delta = 0$, which is expressed in terms of product moments as

$$LR^{-2/T}(\mathcal{M}_3|\mathcal{M}_1) = \frac{|S_{00\cdot a,v}|}{|S_{00\cdot v}|} = \frac{|S_{\varepsilon\varepsilon\cdot a,v}|}{|S_{\varepsilon\varepsilon\cdot v}|} = \frac{|S_{\varepsilon\varepsilon\cdot v} - S_{\varepsilon a\cdot v}S_{aa\cdot v}^{-1}S_{a\varepsilon\cdot v}|}{|S_{\varepsilon\varepsilon\cdot v}|},$$

which is just another expression for (31), but given in terms of processes that are normalized. Hence with $N = TS_{\varepsilon\varepsilon\cdot v}^{-1}S_{\varepsilon a\cdot v}S_{aa\cdot v}^{-1}S_{a\varepsilon\cdot v}$, which is $O_P(1)$, we find

$$-2\log LR(\mathcal{M}_3|\mathcal{M}_1) = -T\log|I_n - T^{-1}N| \stackrel{1}{=} \operatorname{tr}\{N\} + \frac{1}{2T}\operatorname{tr}\{N^2\}, \tag{37}$$

indicating that we have kept terms of order $O_P(T^{-1})$.

THEOREM 2: *The likelihood ratio test for cointegrating rank in model* (1) *has the expansion*

$$\begin{aligned}
&-2\log LR(\mathcal{M}_2|\mathcal{M}_1)\\
&\stackrel{1}{=} T\operatorname{tr}\{S_{ab\cdot v}S_{bb\cdot v}^{-1}S_{ba\cdot v}S_{aa\cdot v}^{-1}\} + \frac{1}{2}T\operatorname{tr}\{(S_{ab\cdot v}S_{bb\cdot v}^{-1}S_{ba\cdot v}S_{aa\cdot v}^{-1})^2\}\\
&\quad - 2T\operatorname{tr}\{S_{bb}^{-1}S_{ba\cdot v}S_{aa}^{-1}S_{au\cdot v,b}\kappa S_{vv}^{-1}S_{vb}\} - \operatorname{tr}\{\kappa S_{vv}^{-1}\kappa S_{ua}S_{aa}^{-1}S_{ab}S_{ba}S_{aa}^{-1}S_{au}\}\\
&\quad - \operatorname{tr}\{S_{ba}S_{aa}^{-1}S_{ab}S_{bv}S_{vv}^{-1}\kappa^2 S_{vv}^{-1}S_{vb}\} + \operatorname{tr}\{S_{ua}S_{aa}^{-1}S_{au}\kappa S_{vv}^{-1}S_{vb}S_{bv}S_{vv}^{-1}\kappa\}\\
&\quad + 2\operatorname{tr}\{S_{ba}S_{aa}^{-1}S_{au}\kappa S_{vv}^{-1}S_{vu}\kappa S_{vv}^{-1}S_{vb}\},
\end{aligned}$$

where $\kappa = (\alpha'\Omega^{-1}\alpha)^{-\frac{1}{2}}$.

PROOF: We start by expanding the matrix N (see (37)) by introducing the variables U_t and B_t (see (10)) and finding

$$\begin{aligned}
\operatorname{tr}\{N\} &= T\operatorname{tr}\{S_{a\varepsilon\cdot v}S_{\varepsilon\varepsilon\cdot v}^{-1}S_{\varepsilon a\cdot v}S_{aa\cdot v}^{-1}\}\\
&= T\operatorname{tr}\left\{\begin{pmatrix} S_{ua\cdot v}\\ S_{ba\cdot v}\end{pmatrix}'\begin{pmatrix} S_{uu\cdot v} & S_{ub\cdot v}\\ S_{bu\cdot v} & S_{bb\cdot v}\end{pmatrix}^{-1}\begin{pmatrix} S_{ua\cdot v}\\ S_{ba\cdot v}\end{pmatrix}S_{aa\cdot v}^{-1}\right\}\\
&= T\operatorname{tr}\{S_{ab\cdot v}S_{bb\cdot v}^{-1}S_{ba\cdot v}S_{aa\cdot v}^{-1}\} + T\operatorname{tr}\{S_{au\cdot v,b}S_{uu\cdot v,b}^{-1}S_{ua\cdot v,b}S_{aa\cdot v}^{-1}\}\\
&= \operatorname{tr}\{N_1 + N_2\},
\end{aligned}$$

where N_1 and N_2 are $O_P(1)$. From (37) we find

$$-2\log LR(\mathcal{M}_3|\mathcal{M}_1) \stackrel{1}{=} \operatorname{tr}\{N_1 + N_2\} + \frac{1}{2}T^{-1}\operatorname{tr}\{(N_1 + N_2)^2\}. \tag{38}$$

The term N_2 can be rewritten as follows

$$\begin{aligned}
N_2 &= TS_{au\cdot v,b}S_{uu\cdot v,b}^{-1}S_{ua\cdot v,b}S_{aa\cdot v}^{-1}\\
&= TS_{au\cdot v,b}S_{uu\cdot v,b}^{-1}S_{ua\cdot v,b}S_{aa\cdot v,b}^{-1} + TS_{au\cdot v,b}S_{uu\cdot v,b}^{-1}S_{ua\cdot v,b}S_{aa\cdot v}^{-1}(I_{n_a} - S_{aa\cdot v}S_{aa\cdot v,b}^{-1})\\
&= N_{21} + N_2(I_{n_a} - S_{aa\cdot v}S_{aa\cdot v,b}^{-1}) = N_{21} + N_{22},
\end{aligned} \tag{39}$$

where $S_{aa\cdot v,b}S_{aa\cdot v}^{-1} = (S_{aa\cdot v} - S_{ab\cdot v}S_{bb\cdot v}^{-1}S_{ba\cdot v})S_{aa\cdot v}^{-1} = I_{n_a} - T^{-1}N_1$, so that

$$S_{aa\cdot v}S_{aa\cdot v,b}^{-1} = (I_{n_a} - T^{-1}N_1)^{-1} \stackrel{1}{=} I_{n_a} - T^{-1}N_1.$$

Hence the term N_{22} in (39), becomes

$$N_{22} = N_2\big(I_{n_a} - \big(I_{n_a} - T^{-1}N_1\big)^{-1}\big) \overset{1}{=} -T^{-1}N_2N_1,$$

which cancels the double product from $\frac{1}{2}T^{-1}(N_1+N_2)^2$ in (38). We next find $N_2^2 \overset{0}{=} (S_{au}S_{ua}S_{aa}^{-1})^2$, where we keep only terms of order $O_P(1)$. Thus

$$\begin{aligned}(40)\qquad & -2\log LR(\mathscr{M}_3|\mathscr{M}_1)\\ &\overset{1}{=} \operatorname{tr}\{N_1\} + \operatorname{tr}\{N_{21}\} + \frac{1}{2}T^{-1}\operatorname{tr}\{N_1^2\} + \frac{1}{2}T^{-1}\operatorname{tr}\{N_2^2\}\\ &\overset{1}{=} T\operatorname{tr}\{S_{ab\cdot v}S_{bb\cdot v}^{-1}S_{ba\cdot v}S_{aa\cdot v}^{-1}\} + T\operatorname{tr}\{S_{aa\cdot v,b}^{-1}S_{au\cdot v,b}S_{uu\cdot v,b}^{-1}S_{ua\cdot v,b}\}\\ &\quad + \frac{1}{2}T\operatorname{tr}\{(S_{ab\cdot v}S_{bb\cdot v}^{-1}S_{ba\cdot v}S_{aa\cdot v}^{-1})^2\} + \frac{1}{2}T^{-1}\operatorname{tr}\{(S_{au}S_{ua}S_{aa}^{-1})^2\}.\end{aligned}$$

In Johansen (2000, Theorem 3) it is shown that

$$\begin{aligned}& -2\log LR(\mathscr{M}_3|\mathscr{M}_2)\\ &\overset{1}{=} T\operatorname{tr}\{S_{aa\cdot v,b}^{-1}S_{au\cdot v,b}S_{uu\cdot v,b}^{-1}S_{ua\cdot v,b}\} + \frac{1}{2}T^{-1}\operatorname{tr}\{(S_{ua}S_{aa}^{-1}S_{au})^2\}\\ &\quad + 2T\operatorname{tr}\{S_{aa}^{-1}S_{ua\cdot v,b}\kappa S_{vv}^{-1}S_{vb}S_{bb}^{-1}S_{ba\cdot v}\} + \operatorname{tr}\{\kappa S_{vv}^{-1}\kappa S_{ua}S_{aa}^{-1}S_{ab}S_{ba}S_{aa}^{-1}S_{au}\}\\ &\quad + \operatorname{tr}\{S_{ba}S_{aa}^{-1}S_{ab}S_{bv}S_{vv}^{-1}\kappa^2S_{vv}^{-1}S_{vb}\} - \operatorname{tr}\{S_{ua}S_{aa}^{-1}S_{au}\kappa S_{vv}^{-1}S_{vb}S_{bv}S_{vv}^{-1}\kappa\}\\ &\quad - 2\operatorname{tr}\{S_{ba}S_{aa}^{-1}S_{au}\kappa S_{vv}^{-1}S_{vu}\kappa S_{vv}^{-1}S_{vb}\}.\end{aligned}$$

Subtracting this result from (40), we have finished the proof of the expansion in Theorem 2. *Q.E.D.*

A.4. *The Moments and Their Asymptotic Properties*

LEMMA 1: *The moments* (19) *are invariant to any full rank linear transformation of* A_t, *and for any such moment* $\widetilde{M}$ *we have*

$$(41)\qquad E[\widetilde{M}] = \operatorname{tr}\{E[\widetilde{M}]\}n_b^{-1}I_{n_b}.$$

PROOF: The process A_{t-1} enters the moments through the factor

$$(42)\qquad A'_{t-1}\left(\sum_{u=1}^{T}A_{u-1}A'_{u-1}\right)^{-1}A_{s-1},$$

which is clearly invariant under nonsingular linear transformations of A_{t-1}. For any orthogonal matrix $O(n_b\times n_b)$ $(O'O = I_{n_b})$ define $K_O = \operatorname{diag}\{O', 1\}$. Replacing B_t by $O'B_t$, replacing A_{t-1} by K_OA_{t-1}, which by (42) does not change $\widetilde{M}$, and we have

$$(43)\qquad \widetilde{M}(O'B) = O'\widetilde{M}O.$$

Taking expectations, we find that $E[\widetilde{M}]$ is invariant under all orthogonal transformations and hence proportional to I_{n_b}. This proves (41). *Q.E.D.*

In order to find the asymptotic properties of the moments we define the processes

$$(44)\qquad T^{-\frac{1}{2}}\sum_{t=1}^{[Ts]} B_t \xrightarrow{w} W^b(s) = W(s),$$

$$T^{-\frac{1}{2}}\sum_{t=1}^{[Ts]} U_t \xrightarrow{w} W^u(s),$$

$$T^{-\frac{1}{2}}\sum_{t=1}^{[Ts]} (B_{t-i}B_t' - \delta_{i0}I_{n_b}) \xrightarrow{w} W_i^{bb}(s) \qquad (i = 0, 1, \dots),$$

$$T^{-\frac{1}{2}}\sum_{t=1}^{[Ts]} B_t U_{t-i}' \xrightarrow{w} W_i^{bu}(s) \qquad (i = 0, 1, \dots),$$

$$T^{-\frac{1}{2}}\sum_{t=1}^{[Ts]} (U_t U_{t-i}' - \delta_{i0}I_{n_u}) \xrightarrow{w} W_i^{uu}(s) \qquad (i = 0, 1, \dots).$$

It is seen that all these processes are mutually independent. Suitably normalized the process $A_{[Ts]}$ (see (33)) converges to a limit F, which depends on W and the deterministic terms. The limit of the moments can be expressed in terms of W, W_i^{bb}, and F.

LEMMA 2: *Let* A_{t-1} *be given by* (33); *then normalizing by* $D_T = \text{diag}\{T^{-1/2}I_{n_b}, T^{-n_d}\}$ *we find*

$$(45)\qquad D_T A_{[Ts]-1} = \left(\begin{matrix} T^{-\frac{1}{2}}\sum_{i=1}^{[Ts]-1} B_i \\ ([Ts]/T)^{n_d} \end{matrix}\middle| d_t \right) \xrightarrow{w} F(s) = \left(\begin{matrix} W(s) \\ s^{n_d} \end{matrix}\middle| 1, \dots, s^{n_d-1}\right).$$

It follows (*see* (19)) *that*

$$(46)\qquad M \xrightarrow{w} \int_0^1 (dW)F'\left(\int_0^1 FF'\,ds\right)^{-1}\int_0^1 F(dW)',$$

$$M_+ \xrightarrow{w} \int_0^1 (dW)F'\left(\int_0^1 FF'\,ds\right)^{-1}\left(\int_0^1 F(dW)' + (I_{n_b}, 0)'\right),$$

$$M_{++} \xrightarrow{w} \left((I_{n_b}, 0) + \int_0^1 (dW)F'\right)\left(\int_0^1 FF'\,ds\right)^{-1}\left(\int_0^1 F(dW)' + (I_{n_b}, 0)'\right),$$

$$T^{-1/2}M_- \xrightarrow{w} W_1^{bb}\int_0^1 (dW)F'\left(\int_0^1 FF'\,ds\right)^{-1}\left(\int_0^1 F(dW)' + (I_{n_b}, 0)'\right),$$

$$T^{-1/2}K(j) \xrightarrow{w} (W_{j+1}^{bb} - W_1^{bb})\int_0^1 (dW)F'\left(\int_0^1 FF'\,ds\right)^{-1}\left(\int_0^1 F(dW)' + (I_{n_b}, 0)'\right).$$

Note that the limits of $T^{-1/2}M_-$ and $T^{-1/2}K(j)$ have expectation zero because W_i^{bb} is independent of W.

PROOF: The results (45) and (46) follow from standard results about Brownian motion. *Q.E.D.*

In the calculations in the proof of Theorem 1 we often replace $\sum_{t=1}^T A_{t-1}A_{t-1-k}'$ by M_{aa} since it holds that

$$\left(\sum_{t=1}^T A_{t-1}A_{t-1}'\right)^{-1}\sum_{t=1}^T A_{t-1}A_{t-1-k}' \xrightarrow{P} I_{n_a}, \quad \text{for all } k,$$

and similarly we replace $\sum_{t=1}^T A_{t-1}B_{t-1-k}'$ by $M_{ab}^+ = \sum_{t=1}^T A_{t-1}B_{t-1}'$, if $k \geq 0$ and by M_{ab}, for $k < 0$.

LEMMA 3: *The variable* $T^{-1/2}(\lambda' M_{yb}\nu, \mu'(M_{yy}-T\Sigma)\xi)$ *is asymptotically Gaussian with mean zero and variance matrix given by*

$$\text{var}[T^{-1/2}\lambda' M_{yb}\nu] \to \lambda'\Sigma\lambda\nu'\nu, \tag{47}$$

$$\text{cov}\big[T^{-1/2}\lambda' M_{yb}\nu, T^{-1/2}\mu'(M_{yy}-T\Sigma)\xi\big]$$

$$\longrightarrow \sum_{m=0}^{\infty}\lambda'\gamma(m+1)\mu\nu'\psi_m'\xi + \lambda'\gamma(m+1)\xi\nu'\psi_m'\mu,$$

$$\text{var}\big[T^{-1/2}\mu'\big(M_{yy}-T\Sigma\big)\xi\big] \longrightarrow \sum_{m=-\infty}^{\infty}\mu'\gamma(m)\mu\xi'\gamma(m)\xi + \mu'\gamma(m)\xi\xi'\gamma(m)\mu.$$

It follows that

$$T^{-1}E\big[M_{by}\Sigma^{-1}M_{yb}\big] \longrightarrow n_y I_{n_b}, \quad T^{-1}E[M_{yb}M_{by}] \longrightarrow n_b\Sigma^{-1}, \tag{48}$$

$$E\big[\big(T^{-1}M_{yy}-\Sigma\big)\Sigma^{-1}M_{yb}\big] \longrightarrow \sum_{k=0}^{\infty}\gamma(k+1)'\Sigma^{-1}\psi_k + \psi_k\text{tr}\big\{\Sigma^{-1}\gamma(k+1)\big\}.$$

Conditional on $B=\{B_t\}$ *the distribution of* $M_{aa}^{-1/2}M_{ay}$ *is Gaussian with the same limit distribution as*

$$N_{n_a\times n_y}(M_{aa}^{-1/2}M_{ab}^{+}\psi', I_{n_a}\otimes\theta\theta'). \tag{49}$$

It follows that

$$E\big[M_{aa}^{-1/2}M_{ay}\Sigma^{-1}M_{ya}M_{aa}^{-1/2}|B\big] \overset{0}{=} M_{aa}^{-1/2}M_{ab}^{+}\psi'\Sigma^{-1}\psi M_{ba}^{+}M_{aa}^{-1/2} + I_{n_a}\text{tr}\{\theta'\Sigma^{-1}\theta\}, \tag{50}$$

$$E\big[M_{ya}M_{aa}^{-1}M_{ay}|B\big] \overset{0}{=} \psi M_{ba}^{+}M_{aa}^{-1}M_{ab}^{+}\psi' + n_a\theta\theta'. \tag{51}$$

Finally $T^{-1/2}(\lambda' M_{yb}\nu, \mu'(M_{yy}-T\Sigma)\xi)$ *is asymptotically independent of* $M_{aa}^{-1/2}M_{ay}$.

PROOF: The Central Limit Theorem shows that $T^{-1/2}(\lambda' M_{yb}\nu, \mu'(M_{yy}-T\Sigma)\xi)$ is asymptotically Gaussian with mean zero. We apply in the following the well known formula for the moments of a multivariate Gaussian distribution X with mean zero and variance Σ: $\text{cov}[X_iX_j, X_kX_l] = \Sigma_{ik}\Sigma_{jl} + \Sigma_{il}\Sigma_{jk}$, from which it follows that for any linear processes S_t, H_t, K_t, and L_t with exponentially decreasing coefficients and independent identically distributed Gaussian errors we have

$$T^{-1}\text{cov}[\lambda' M_{sh}\nu, \mu' M_{kl}\xi] \longrightarrow \sum_{m=-\infty}^{\infty}\lambda'\gamma^{sk}(m)\mu\nu'\gamma^{hl}(m)\xi + \lambda'\gamma^{sl}(m)\xi\nu'\gamma^{hk}(m)\mu, \tag{52}$$

where for instance $\gamma^{sk}(m)$ is the covariance function for the processes S_t and K_t.

The formula (52) implies the results in (47) and (48). As an example we find the covariance in (47) and the corresponding moment in (48). Because $\gamma^{by}(m) = \text{cov}(B_t, Y_{t+m}') = 0, m<0$ and ψ_m' for $m \geq 0$ we get

$$\begin{aligned}
&\text{cov}[T^{-1/2}\lambda' M_{yb}\nu, T^{-1/2}\mu'(M_{yy}-T\Sigma)\xi] \\
&= T^{-1}\sum_{t,s}\text{cov}[\lambda' Y_{t-1}B_t'\nu, \mu' Y_{s-1}Y_{s-1}'\xi] \\
&= T^{-1}\sum_{t,s}\big(\lambda'\gamma(s-t)\mu\nu'\gamma^{by}(s-t-1)\xi + \lambda'\gamma(s-t)\xi\nu'\gamma^{by}(t-s-1)\mu\big) \\
&= T^{-1}\sum_{m=-T}^{T}(T-|m|)(\lambda'\gamma(m)\mu\nu'\gamma^{by}(m-1)\xi + \lambda'\gamma(m)\xi\nu'\gamma^{by}(m-1)\mu) \\
&\longrightarrow \sum_{m=0}^{\infty}\lambda'\gamma(m+1)\mu\nu'\psi_m'\xi + \lambda'\gamma(m+1)\xi\nu'\psi_m'\mu,
\end{aligned}$$

which is the limiting covariance in (47).

From this it follows that for $\Sigma^{-1} = \sum_{k=1}^{n_y} \rho_k \nu_k \nu_k'$ we find the moment

$$\begin{aligned}
&E[\mu'(T^{-1}M_{yy} - \Sigma)\Sigma^{-1}M_{yb}\nu] \\
&\quad = T^{-1}\sum_{k,t,s} \rho_k \operatorname{cov}[\mu' Y_{t-1} Y_{t-1}' \nu_k, \nu_k' Y_{s-1} B_s' \nu] \\
&\quad \longrightarrow \sum_k \rho_k \sum_{m=0}^{\infty} \nu_k' \gamma(m+1)\mu\nu'\psi_m'\nu_k + \nu_k'\gamma(m+1)\nu_k\nu'\psi_m'\mu \\
&\quad = \sum_k \rho_k \sum_{m=0}^{\infty} \nu'\psi_m'\nu_k\nu_k'\gamma(m+1)\mu + \operatorname{tr}\{\nu_k\nu_k'\gamma(m+1)\}\nu'\psi_m'\mu \\
&\quad = \sum_{m=0}^{\infty} \nu'\psi_m'\Sigma^{-1}\gamma(m+1)\mu + \operatorname{tr}\{\Sigma^{-1}\gamma(m+1)\}\nu'\psi_m'\mu \\
&\quad = \sum_{m=0}^{\infty} \mu'\gamma(m+1)'\Sigma^{-1}\psi_m\nu + \operatorname{tr}\{\Sigma^{-1}\gamma(m+1)\}\mu'\psi_m\nu,
\end{aligned}$$

which is the second expression in (48). To show (49) we find that

$$M_{aa}^{-1/2}M_{ay} = M_{aa}^{-1/2}\sum_{t,i} A_{t-1}(U_{t-i-1}'\theta_i' + B_{t-i-1}'\psi_i') \tag{53}$$

conditional on the variables $B = \{B_t\}$ (and hence $A = \{A_t\}$) has a Gaussian distribution with mean

$$E[M_{aa}^{-1/2}M_{ay}|B] = M_{aa}^{-1/2}\sum_{t,i} A_{t-1}B_{t-i-1}'\psi_i' \overset{0}{=} M_{aa}^{-1/2}M_{ab}^{+}\psi'.$$

The variance can be found from

$$\operatorname{var}[\lambda' M_{aa}^{-1/2}M_{ay}\mu|B] = \lambda' M_{aa}^{-1/2}\sum_{t,s,i,j} A_{t-1}E[U_{t-i-1}'\theta_i'\mu\mu'\theta_j U_{s-j-1}]A_{s-1}'M_{aa}^{-1/2}\lambda.$$

We only get a contribution for $t-i = s-j$, and for $M_{aa}^{-1/2}\sum_t A_{t-1}A_{t-i+j-1}'M_{aa}^{-1/2} \overset{0}{=} I_{n_a}$ we find

$$\operatorname{var}\left[\lambda' M_{aa}^{-1/2}M_{ay}\mu|B\right] \overset{0}{=} \lambda'\lambda\sum_{i,j}\operatorname{tr}\{\theta_i'\mu\mu'\theta_j\} = \lambda'\lambda\mu'\theta\theta'\mu.$$

This proves (49). The expressions for $E[M_{aa}^{-1/2}M_{ay}\Sigma^{-1}M_{ya}M_{aa}^{-1/2}|B]$ and $E[M_{ya}M_{aa}^{-1}M_{ay}|B]$ are found as above. Finally it follows from (53) that the limit distribution is determined by the Brownian motions W^u and W^b, whereas similar expressions for $T^{-1/2}(M_{yy} - TI_{n_y})$ and $T^{-1/2}M_{yb}$ show that their limits are determined by $(W_i^{uu}, W_i^{ub}, W_i^{bb}, i = 0, \dots)$, which are independent of W^u and W^b; see (44). *Q.E.D.*

LEMMA 4: *The moments M_{bd} are independent of $M_{ba}M_{aa}^{-1}M_{ab}$ and $M_{ba}M_{aa}^{-1}M_{ab}^{+}$.*

PROOF: We construct a statistical model in which $M_{ba}M_{aa}^{-1}M_{ab}$ and $M_{ba}M_{aa}^{-1}M_{ab}^{+}$ are ancillary and M_{bd} is minimal sufficient and boundedly complete so that they are independent by Basu's theorem; see Basu (1955).

If $n_d = 0$, then $d_t = 0$ and $M_{bd} = 0$, and there is nothing to prove. For $n_d \geq 1$ we consider the statistical model for the n_b dimensional process $\widetilde{B}_t$ with parameters $\Phi : \widetilde{B}_t = \Phi d_t + B_t$, $t = 1, \ldots, T$, where B_t are i.i.d. $N_{n_b}(0, I_{n_b})$. We define the statistic

$$M_{\tilde{b}d} = \sum_{t=0}^{T} \widetilde{B}_t d_t' = \sum_{t=0}^{T} \Phi d_t d_t' + \sum_{t=0}^{T} B_t d_t' = \Phi M_{dd} + M_{bd},$$

which is distributed as $N_{n_b \times n_d}(\Phi M_{dd}, I_{n_b} \otimes M_{dd})$, which is boundedly complete. Let

$$\widetilde{A}_{t-1} = \left(\begin{array}{c|c} \sum_{i=0}^{t-1} \widetilde{B}_i & \\ & d_t \\ t^{n_d} & \end{array} \right) = \begin{pmatrix} I_{n_b} & \Phi_{n_d-1}/n_d \\ 0 & 1 \end{pmatrix} \left(\begin{array}{c|c} \sum_{i=0}^{t-1} B_i & \\ & d_t \\ t^{n_d} & \end{array} \right).$$

We define the functional $\widetilde{M} = M_{\tilde{b}\tilde{a}} M_{\tilde{a}\tilde{a}}^{-1} M_{\tilde{a}\tilde{b}}$ and note that by the invariance under multiplication by nonsingular matrices we have $\widetilde{M} = M_{\tilde{b}\tilde{a}} M_{\tilde{a}\tilde{a}}^{-1} M_{\tilde{a}\tilde{b}} = M_{\tilde{b}a} M_{aa}^{-1} M_{a\tilde{b}}$ but we also have that

$$M_{\tilde{b}a} = \sum_t \widetilde{B}_t A_{t-1}' = \sum_t (\Phi d_t + B_t) A_{t-1}' = \sum_t B_t A_{t-1}',$$

since $M_{ad} = 0$. Thus $\widetilde{M} = M$, and similarly $\widetilde{M}_+ = M_+$. Therefore the statistics M and M_+ are ancillary and hence independent of M_{bd} by Basu's theorem. *Q.E.D.*

A.5. *Proof of Theorem 1*

PROOF: We apply the expansion from Theorem 2, and we find that the expectation of the last five terms are calculated in Johansen (2000, proof of Theorem 4, p. 775). It is shown there that the total contribution of these terms is

$$-(n-r)(n-r+1)v_\theta, \tag{54}$$

where $v_\theta = \mathrm{tr}\{\theta' \Sigma^{-1} \theta\}$; see (14). What remains are the two terms

$$\begin{aligned} TE\big[\mathrm{tr}\{S_{ab\cdot v} S_{bb\cdot v}^{-1} S_{ba\cdot v} S_{aa\cdot v}^{-1}\}\big] &+ \frac{1}{2} TE\big[\mathrm{tr}\{(S_{ab\cdot v} S_{bb\cdot v}^{-1} S_{ba\cdot v} S_{aa\cdot v}^{-1})^2\}\big] \\ &= E\big[\mathrm{tr}\{S_{ab\cdot v} (T^{-1} S_{bb\cdot v})^{-1} S_{ba\cdot v} S_{aa\cdot v}^{-1}\}\big] \\ &\quad + \frac{1}{2} T^{-1} E\big[\mathrm{tr}\{(S_{ab\cdot v} (T^{-1} S_{bb\cdot v})^{-1} S_{ba\cdot v} S_{aa\cdot v}^{-1})^2\}\big]. \end{aligned} \tag{55}$$

We start with the second term in (55). The factor T^{-1} in front means that we can replace matrices by their limit. From the definition of $S_{aa\cdot v}$ (see (36)) we find with $Y_t = (V_t', Z_t')'$ that because A_{t-1} is orthogonalized on d_t we have $M_{aa\cdot d} = M_{aa}$ and $M_{ya\cdot d} = M_{ya}$ and then get

$$\begin{aligned} M_{aa}^{-1/2} S_{aa\cdot v} M_{aa}^{-1/2} &= M_{aa}^{-1/2} M_{aa\cdot v, z, d} M_{aa}^{-1/2} = M_{aa}^{-1/2} M_{aa\cdot y, d} M_{aa}^{-1/2} \\ &= I_{n_a} - M_{aa}^{-1/2} M_{ay\cdot d} M_{yy\cdot d}^{-1} M_{ya\cdot d} M_{aa}^{-1/2} \\ &= I_{n_a} - T^{-1} M_{aa}^{-1/2} M_{ay} (T^{-1} M M_{yy\cdot d})^{-1} M_{ya} M_{aa}^{-1/2} \\ &= I_{n_a} + O_P(T^{-1}). \end{aligned}$$

Similarly we find the evaluations

$$\begin{aligned} M_{aa}^{-1/2} S_{ab\cdot v} &= M_{aa}^{-1/2} M_{ab\cdot y, d} \\ &= M_{aa}^{-1/2} M_{ab} - T^{-1/2} M_{aa}^{-1/2} M_{ay\cdot d} (T^{-1} M_{yy\cdot d})^{-1} T^{-1/2} M_{yb\cdot d} \\ &= M_{aa}^{-1/2} M_{ab} + O_P(T^{-1/2}), \end{aligned}$$

$$\begin{aligned} T^{-1}S_{bb\cdot v} &= T^{-1}M_{bb\cdot y,\,d} \\ &= T^{-1}M_{bb\cdot d} - T^{-1}(T^{-1/2}M_{by\cdot d}(T^{-1}M_{yy\cdot d})^{-1}T^{-1/2}M_{yb\cdot d}) \\ &= T^{-1}M_{bb\cdot d} + O_P(T^{-1}), \end{aligned}$$

so that

$$\left(S_{ab\cdot v}(T^{-1}S_{bb\cdot v})^{-1}S_{ba\cdot v}S_{aa\cdot v}^{-1}\right)^2 \overset{0}{=} \left((T^{-1}M_{bb\cdot d})^{-1}M_{ba}M_{aa}^{-1}M_{ab}\right)^2,$$

in the sense that the difference tends to zero in probability. We now replace the expectation of the right-hand side by the expectation of the left-hand side and write therefore

$$E\left[\mathrm{tr}\left\{\left(S_{ab\cdot v}(T^{-1}S_{bb\cdot v})^{-1}S_{ba\cdot v}S_{aa\cdot v}^{-1}\right)^2\right\}\right] \overset{0}{=} E\left[\mathrm{tr}\left\{\left((T^{-1}M_{bb\cdot d})^{-1}M_{ba}M_{aa}^{-1}M_{ab}\right)^2\right\}\right]. \tag{56}$$

What remains is the first term of (55). We expand it as follows:

$$\begin{aligned} S_{ab\cdot v}(T^{-1}S_{bb\cdot v})^{-1}S_{ba\cdot v}S_{aa\cdot v}^{-1} &= M_{ab\cdot y,\,d}(T^{-1}M_{bb\cdot y,\,d})^{-1}M_{ba\cdot y,\,d}M_{aa\cdot y,\,d}^{-1} \\ &= \left(M_{ab} - M_{ay}M_{yy\cdot d}^{-1}M_{yb\cdot d}\right)\left(T^{-1}M_{bb\cdot d} - T^{-1}M_{by\cdot d}M_{yy\cdot d}^{-1}M_{yb\cdot d}\right)^{-1} \\ &\quad \times \left(M_{ba} - M_{by\cdot d}M_{yy\cdot d}^{-1}M_{ya}\right)\left(M_{aa} - M_{ay}M_{yy\cdot d}^{-1}M_{ya}\right)^{-1}, \end{aligned}$$

where we have used that $M_{ad} = 0$, since A_{t-1} has been orthogonalized on d_t. We get a number of different terms when expanding and keeping terms of order T^{-1}:

$$\begin{aligned} &S_{ab\cdot v}(T^{-1}S_{bb\cdot v})^{-1}S_{ba\cdot v}S_{aa\cdot v}^{-1} \\ &\quad \overset{1}{=} M_{ab}(T^{-1}M_{bb\cdot d})^{-1}M_{ba}M_{aa}^{-1} \\ &\qquad - \left(M_{ay}M_{yy\cdot d}^{-1}M_{yb\cdot d}\right)(T^{-1}M_{bb\cdot d})^{-1}M_{ba}M_{aa}^{-1} \\ &\qquad - M_{ab}(T^{-1}M_{bb\cdot d})^{-1}\left(M_{by\cdot d}M_{yy\cdot d}^{-1}M_{ya}\right)M_{aa}^{-1} \\ &\qquad + T^{-1}M_{ab}(T^{-1}M_{bb\cdot d})^{-1}\left(M_{by\cdot d}M_{yy\cdot d}^{-1}M_{yb\cdot d}\right)(T^{-1}M_{bb\cdot d})^{-1}M_{ba}M_{aa}^{-1} \\ &\qquad + T^{-1}M_{ab}(T^{-1}M_{bb\cdot d})^{-1}M_{ba}M_{aa}^{-1}\left(M_{ay}(T^{-1}M_{yy\cdot d})^{-1}M_{ya}\right)M_{aa}^{-1} \\ &\qquad + T^{-1}\left(M_{ay}M_{yy\cdot d}^{-1}M_{yb\cdot d}\right)(T^{-1}M_{bb\cdot d})^{-1}\left(M_{by\cdot d}(T^{-1}M_{yy\cdot d})^{-1}M_{ya}\right)M_{aa}^{-1} \\ &\quad = K_0 + T^{-1}(K_1 + K_2 + K_3 + K_4). \end{aligned} \tag{57}$$

The expectation of the trace of the first term is

$$E[\mathrm{tr}\{K_0\}] = E\left[\mathrm{tr}\left\{M_{ab}(T^{-1}M_{bb\cdot d})^{-1}M_{ba}M_{aa}^{-1}\right\}\right],$$

which we combine with (56) to the term

$$\begin{aligned} &TE\left[\mathrm{tr}\left\{M_{ab}M_{bb\cdot d}^{-1}M_{ba}M_{aa}^{-1}\right\}\right] + \frac{1}{2}TE\left[\mathrm{tr}\left\{\left(M_{bb\cdot d}^{-1}M_{ba}M_{aa}^{-1}M_{ab}\right)^2\right\}\right] \\ &\quad \overset{1}{=} -TE\left[\log\left|I_{n_b} - M_{bb\cdot d}^{-1}M_{ba}M_{aa}^{-1}M_{ab}\right|\right], \end{aligned}$$

which is the main term in Theorem 1.

For the remaining terms in (57) we can prove

$$
\begin{aligned}
E[\mathrm{tr}\{K_1\}] &= -2E\big[\mathrm{tr}\big\{\big(M_{ay}(T^{-1}M_{yy\cdot d})^{-1}M_{yb\cdot d}\big)(T^{-1}M_{bb\cdot d})^{-1}M_{ba}M_{aa}^{-1}\big\}\big]\\
&\overset{0}{=} \mathrm{tr}\{E[M]\}(-v_\theta + v_\psi - n_y) + \mathrm{tr}\{E[M_+]\}(-v_\theta - v_\psi + n_y)\\
&\quad - 2\mathrm{tr}\Bigg\{\psi'\Sigma^{-1}\sum_{j=0}^{\infty}\psi_j E[K(j)] + \psi'\Sigma^{-1}\psi E[M_-]\\
&\quad - n_d E[M_+]\psi'\Sigma^{-1}\psi - E[M_+]V\Bigg\},\\
E[\mathrm{tr}\{K_2\}] &= E\big[\mathrm{tr}\big\{M_{ab}(T^{-1}M_{bb\cdot d})^{-1}\big(M_{by\cdot d}M_{yy\cdot d}^{-1}M_{yb\cdot d}\big)\\
&\qquad \times (T^{-1}M_{bb\cdot d})^{-1}M_{ba}M_{aa}^{-1}\big\}\big]\\
&\overset{0}{=} (r + (k-1)n)\mathrm{tr}\{E[M]\},\\
E[\mathrm{tr}\{K_3\}] &= E\big[\mathrm{tr}\big\{M_{ab}(T^{-1}M_{bb\cdot d})^{-1}M_{ba}M_{aa}^{-1}\\
&\qquad \times \big(M_{ay}(T^{-1}M_{yy\cdot d})^{-1}M_{ya}\big)M_{aa}^{-1}\big\}\big]\\
&\overset{0}{=} \mathrm{tr}\{E[M]\}v_\theta + \mathrm{tr}\big\{E[M_+'M_+]\psi'\Sigma^{-1}\psi\big\},\\
E[\mathrm{tr}\{K_4\}] &= T^{-1}E\big[\mathrm{tr}\big\{\big(M_{ay}(T^{-1}M_{yy\cdot d})^{-1}M_{yb\cdot d}\big)(T^{-1}M_{bb\cdot d})^{-1}\\
&\qquad \times \big(M_{by\cdot d}(T^{-1}M_{yy\cdot d})^{-1}M_{ya}\big)M_{aa}^{-1}\big\}\big]\\
&\overset{0}{=} (n-r+1)(n-r)v_\theta + (n-r)\mathrm{tr}\{E[M_{++}]\psi'\Sigma^{-1}\psi\},
\end{aligned}
$$

where $\psi'\Sigma^{-1}\psi$ is given by (14) and V by (15). Adding these contributions we note that the first term in $E[\mathrm{tr}\{K_4\}]$ cancels (54) and we find the result in Theorem 1.

The final part of this proof contains an evaluation of $E[\mathrm{tr}\{K_1\}], \ldots, E[\mathrm{tr}\{K_4\}]$, in order to prove these relations. Inspecting the expressions for $K_1, \ldots, K_4$, one can see that one can normalize Y_t so that $\Sigma = I_{n_y}$, and A_{t-1} so that $M_{aa} = I_{n_a}$. This simplifies the notation.

K1: If $\Sigma = I_{n_y}$, then

$$
\begin{aligned}
V_\theta &= \theta\theta' = \sum_{h=-\infty}^{\infty}\gamma_\theta(h) = \sum_{h=0}^{\infty}\gamma_\theta(h) + \sum_{h=1}^{\infty}\gamma_\theta(h)',\\
V_\psi &= \psi\psi' = \sum_{h=-\infty}^{\infty}\gamma_\psi(h) = \sum_{h=0}^{\infty}\gamma_\psi(h) + \sum_{h=1}^{\infty}\gamma_\psi(h)',
\end{aligned}
$$

so that with $\tau_\theta = \mathrm{tr}\{\gamma_\theta(0)\}$ and $\tau_\psi = \mathrm{tr}\{\gamma_\psi(0)\}$ we find

$$
\begin{aligned}
&2\mathrm{tr}\Bigg\{\sum_{h=1}^{\infty}\gamma_\theta(h)\Bigg\} = \mathrm{tr}\{V_\theta\} - \mathrm{tr}\{\gamma_\theta(0)\} = v_\theta - \tau_\theta,\\
&2\mathrm{tr}\Bigg\{\sum_{h=0}^{\infty}\gamma_\theta(h)\Bigg\} = \mathrm{tr}\{V_\theta\} + \mathrm{tr}\{\gamma_\theta(0)\} = v_\theta + \tau_\theta,\\
&2\mathrm{tr}\Bigg\{\sum_{h=1}^{\infty}\gamma_\psi(h)\Bigg\} = \mathrm{tr}\{V_\psi\} - \mathrm{tr}\{\gamma_\psi(0)\} = v_\psi - \tau_\psi,\\
&2\mathrm{tr}\Bigg\{\sum_{h=0}^{\infty}\gamma_\psi(h)\Bigg\} = \mathrm{tr}\{V_\psi\} + \mathrm{tr}\{\gamma_\psi(0)\} = v_\psi + \tau_\psi.
\end{aligned}
\tag{58}
$$

Note that $\tau_\theta + \tau_\psi = \mathrm{tr}\{\gamma_\theta(0) + \gamma_\psi(0)\} = \mathrm{tr}\{\Sigma\} = n_y$. We find that

$$\begin{aligned}
\mathrm{tr}\{K_1\} &= -2\mathrm{tr}\{M_{ay}(T^{-1}M_{yy\cdot d})^{-1}M_{yb\cdot d}(T^{-1}M_{bb\cdot d})^{-1}M_{ba}\} \\
&= -2\,\mathrm{tr}\{M_{ay}\big(I_{n_y} - \big(I_{n_y} - T^{-1}M_{yy\cdot d}\big)\big)^{-1}\big(M_{yb} - M_{yd}M_{dd}^{-1}M_{db}\big) \\
&\qquad \times (T^{-1}M_{bb\cdot d})^{-1}M_{ba}\} \\
&\overset{0}{=} -2\,\mathrm{tr}\{M_{ay}M_{yb}(T^{-1}M_{bb\cdot d})^{-1}M_{ba} + M_{ay}\big(T^{-1}M_{yy\cdot d} - I_{n_y}\big) \\
&\qquad \times M_{yb}(T^{-1}M_{bb\cdot d})^{-1}M_{ba}\} \\
&\quad + 2\mathrm{tr}\{M_{ay}M_{yd}M_{dd}^{-1}M_{db}(T^{-1}M_{bb\cdot d})^{-1}M_{ba}\} \\
&= K_{11} + K_{12} + K_{13}.
\end{aligned}$$

$\mathbf{K}_{11}$:

$$\begin{aligned}
E[K_{11}] &= -2E\big[\mathrm{tr}\{M_{ay}M_{yb}(T^{-1}M_{bb\cdot d})^{-1}M_{ba}\}\big] \\
&= -2E\Bigg[\mathrm{tr}\Bigg\{\sum_{t,s,i,j} A_{t-1}(U'_{t-i-1}\theta'_i + B'_{t-i-1}\psi'_i) \\
&\qquad \times (\theta_j U_{s-j-1} + \psi_j B_{s-j-1})B'_s(T^{-1}M_{bb\cdot d})^{-1}M_{ba}\Bigg\}\Bigg] \\
&= -2E\Bigg[\mathrm{tr}\Bigg\{\sum_{t,s,i,j} A_{t-1}U'_{t-i-1}\theta'_i\theta_j U_{s-j-1}B'_s(T^{-1}M_{bb\cdot d})^{-1}M_{ba}\Bigg\}\Bigg] \\
&\quad -2E\Bigg[\mathrm{tr}\Bigg\{\sum_{t,s,i,j} A_{t-1}B'_{t-i-1}\psi'_i\psi_j B_{s-j-1}B'_s(T^{-1}M_{bb\cdot d})^{-1}M_{ba}\Bigg\}\Bigg] \\
&= K_{111} + K_{112}.
\end{aligned}$$

Note that the second term is $O_P(T^{1/2})$, but the expectation turns out to be $O(1)$. In the first term we get a contribution for $t-i-1 = s-j-1$ because of the independence of U and B. The term is $O(1)$ and we replace $(T^{-1}M_{bb.d})^{-1}$ by I_{n_b}, and find

$$\begin{aligned}
E[K_{111}] &\overset{0}{=} -2E\Bigg[\mathrm{tr}\Bigg\{\sum_{t,i,j} A_{t-1}B'_{t-i+j}M_{ba}\Bigg\}\Bigg]\mathrm{tr}\{\theta'_i\theta_j\} \\
&\overset{0}{=} -2E[\mathrm{tr}\{M_{ab}M_{ba}\}]\sum_{i\le j}\mathrm{tr}\{\theta'_i\theta_j\} - 2E[\mathrm{tr}\{M^+_{ab}M_{ba}\}]\sum_{i>j}\mathrm{tr}\{\theta'_i\theta_j\},
\end{aligned}$$

where we have replaced $\sum_t A_{t-1}B'_{t-i+j}$ with M_{ab} if $i \le j$ and M^+_{ab} if $i > j$. We then get (see (58))

$$\begin{aligned}
E[K_{111}] &\overset{0}{=} -2\mathrm{tr}\{E[M]\}\sum_{h=0}^{\infty}\mathrm{tr}\{\gamma_\theta(h)\} - 2\mathrm{tr}\{E[M_+]\}\sum_{h=1}^{\infty}\mathrm{tr}\{\gamma_\theta(h)\} \\
&= -\mathrm{tr}\{E[M]\}(v_\theta + \tau_\theta) - \mathrm{tr}\{E[M_+]\}(v_\theta - \tau_\theta).
\end{aligned}$$

In the term K_{112} we write $A_{t-1} = A_{t-i-1} + \sum_{j=t-i}^{t-1}\Delta A_j$ and find

$$\begin{aligned}
E[K_{112}] &= -2E\Bigg[\mathrm{tr}\Bigg\{\sum_{t,s,i,j} A_{t-i-1}B'_{t-i-1}\psi'_i\psi_j B_{s-j-1}B'_s(T^{-1}M_{bb.d})^{-1}M_{ba}\Bigg\}\Bigg] \\
&\quad -2E\Bigg[\mathrm{tr}\Bigg\{\sum_{t,s,i,j}\Bigg(\sum_{m=1}^{i}\Delta A_{t-m}\Bigg)B'_{t-i-1}\psi'_i\psi_j B_{s-j-1}B'_s(T^{-1}M_{bb.d})^{-1}M_{ba}\Bigg\}\Bigg] \\
&= K_{1121} + K_{1122}.
\end{aligned}$$

In K_{1121} we replace $\sum_t A_{t-i-1}B'_{t-i-1}$ by M^+_{ab} and find (see (19))

$$K_{1121} \overset{0}{=} -2E\left[\mathrm{tr}\left\{\psi\sum_j\psi_j\left(\sum_s B_{s-j-1}B'_s\right)(T^{-1}M_{bb\cdot d})^{-1}M_{ba}M^+_{ab}\right\}\right]$$
$$= -2\mathrm{tr}\left\{\psi'\sum_j\psi_j E[K(j)+M_-]\right\}$$
$$= -2\mathrm{tr}\left\{\psi'\sum_j\psi_j E[K(j)]\right\} - 2\mathrm{tr}\{\psi'\psi E[M_-]\}.$$

The terms $K(j)$ and M_- are $O(T^{1/2})$ (see Lemma 2) but the expectation is of the order of 1. For the next term K_{1122} we find a contribution for $t-m=s$ and $t-i-1=s-j-1$, so that $i=m+j$, that is $i>j$. We then get, replacing $(T^{-1}M_{bb\cdot d})^{-1}$ by I_{n_b},

$$E[K_{1122}] = -2E\left[\mathrm{tr}\left\{\sum_{t,s,i,j}\left(\sum_{m=1}^{i}\Delta A_{t-m}\right)B'_{t-i-1}\psi'_i\psi_j B_{s-j-1}\right.\right.$$
$$\left.\left.\times B'_s(T^{-1}M_{bb.d})^{-1}M_{ba}\right\}\right]$$
$$\overset{0}{=} -2E\left[\mathrm{tr}\left\{\sum_{i>j}\sum_s \Delta A_s B'_{s-j-1}\psi'_i\psi_j B_{s-j-1}B'_s M_{ba}\right\}\right]$$
$$\overset{0}{=} -2E\left[\mathrm{tr}\left\{\sum_s \Delta A_s B'_s M_{ba}\right\}\right]\sum_{i>j}\mathrm{tr}\{\psi'_i\psi_j\}$$
$$\overset{0}{=} -2E\left[\mathrm{tr}\{(M^+_{ab}-M_{ab})M_{ba}\}\right]\mathrm{tr}\left\{\sum_{m=1}^{\infty}\gamma_\psi(m)\right\}$$
$$= -(v_\psi-\tau_\psi)(\mathrm{tr}\{E[M_+]\}-\mathrm{tr}\{E[M]\}).$$

Adding the contributions we find, using $\tau_\psi+\tau_\theta=n_y$ (see (58)),

$$E[K_{11}] \overset{0}{=} -2\mathrm{tr}\left\{\psi'\sum_j\psi_j E[K(j)+M_-]\right\} - \mathrm{tr}\{E[M]\}(v_\theta-v_\psi+n_y)$$
$$-\mathrm{tr}\{E[M_+]\}(v_\theta+v_\psi-n_y). \tag{59}$$

$\mathbf{K}_{12}$: We can replace $T^{-1}M_{yy.d}$ by $T^{-1}M_{yy}$ and $(T^{-1}M_{bb.d})^{-1}$ by I_{n_b} and find

$$E[K_{12}] \overset{0}{=} 2E\left[\mathrm{tr}\left\{M_{ay}\left(T^{-1}M_{yy}-I_{n_y}\right)M_{yb}M_{ba}\right\}\right].$$

Because $M_{ba}M_{ay}$ is asymptotically independent of $(T^{-1}M_{yy}-I_{n_y})M_{yb}$ (see Lemma 3), we replace the expectation by the product of the expectations and write

$$E[K_{12}] \overset{0}{=} 2\mathrm{tr}\left\{E\left[\left(T^{-1}M_{yy}-I_{n_y}\right)M_{yb}\right]E[M_{ba}M_{ay}]\right\}.$$

From (48) and (49) we then get that it equals

$$E[K_{12}] \overset{0}{=} 2\mathrm{tr}\left\{\sum_{k=0}^{\infty}(\gamma(k+1)'\psi_k+\psi_k\mathrm{tr}\{\gamma(k+1)'\})E[M_{ba}M^+_{ab}\psi']\right\}$$
$$= 2\mathrm{tr}\{E[M_+]V\}, \tag{60}$$

where V is defined in (15).

$\mathbf{K}_{13}$: Again we replace $(T^{-1}M_{bb.d})^{-1}$ by I_{n_b} and find

$$
\begin{aligned}
E[K_{13}] &= 2E\big[\operatorname{tr}\{M_{ay}M_{yd}M_{dd}^{-1}M_{db}M_{ba}\}\big] \\
&= 2E\Bigg[\operatorname{tr}\Bigg\{\sum_{t,l,i,m} A_{t-1}(U'_{t-i-1}\theta'_i + B'_{t-i-1}\psi'_i)(\theta_m U_{l-m-1} + \psi_m B_{l-m-1}) \\
&\qquad \times d'_l M_{dd}^{-1}M_{db}M_{ba}\Bigg] \\
&= 2E\Bigg[\operatorname{tr}\Bigg\{\sum_{t,l,i,m} A_{t-1}U'_{t-i-1}\theta'_i\theta_m U_{l-m-1}d'_l M_{dd}^{-1}M_{db}M_{ba}\Bigg] \\
&\quad + 2E\Bigg[\operatorname{tr}\Bigg\{\sum_{t,l,i,m} A_{t-1}B'_{t-i-1}\psi'_i\psi_m B_{l-m-1}d'_l M_{dd}^{-1}M_{db}M_{ba}\Bigg] \\
&= K_{131} + K_{132}.
\end{aligned}
$$

For K_{131} we find a contribution for $t-i-1 = l-m-1$:

$$
\begin{aligned}
E[K_{131}] &= 2E\Bigg[\operatorname{tr}\Bigg\{\sum_{t,i,m} A_{t-1}U'_{t-i-1}\theta'_i\theta_m U_{t-i-1}d'_{t-i+m}M_{dd}^{-1}M_{db}M_{ba}\Bigg\}\Bigg] \\
&\overset{0}{=} 2E\Bigg[\operatorname{tr}\Bigg\{\sum_{t,l,i,m} A_{t-1}d'_{t-i+m}M_{dd}^{-1}M_{db}M_{ba}\Bigg\}\operatorname{tr}\{\theta_i\theta_m\}\Bigg] \\
&= 2E\Bigg[\operatorname{tr}\Bigg\{\sum_{t,l,i,m} A_{t-1}d'_t L'^{m-i}M_{dd}^{-1}M_{db}M_{ba}\Bigg\}\operatorname{tr}\{\theta'_i\theta_m\}\Bigg] = 0,
\end{aligned}
$$

since $M_{ad} = 0$. We have used the relation $d'_{t-i+m} = d'_t L'^{m-i}$, for a matrix that is lower triangular with 1 in the diagonal, which holds for $d_t = (1, \ldots, t^{n_d-1})'$ (see (6)). For K_{132} we get

$$
\begin{aligned}
E[K_{132}] &\overset{0}{=} 2E\Bigg[\operatorname{tr}\Bigg\{M_{ab}^{+}\psi' \sum_{l,m}\psi_m B_l d'_{l+m+1}M_{dd}^{-1}M_{db}M_{ba}\Bigg\}\Bigg] \\
&\overset{0}{=} 2E\Bigg[\operatorname{tr}\Bigg\{M_{ab}^{+}\psi' \sum_{m}\psi_m M_{ba}(L')^{m+1}M_{dd}^{-1}M_{db}M_{ba}\Bigg\}\Bigg] \\
&= 2\operatorname{tr}\Bigg\{\psi' \sum_{m}\psi_m E\big[M_{bd}(L')^{m+1}M_{dd}^{-1}M_{db}\big]E\big[M_{ba}M_{ab}^{+}\big]\Bigg\},
\end{aligned}
$$

because M_{bd} is independent of $M_{ba}M_{ab}^{+}$ (see Lemma 4). We therefore evaluate

$$
\begin{aligned}
E\big[M_{ba}(L')^{m+1}M_{dd}^{-1}M_{db}\big] &= \sum_t E\big[B_t d'_t (L')^{m+1}M_{dd}^{-1}d_t B'_t\big] \\
&= \sum_t d'_t (L')^{m+1}M_{dd}^{-1}d_t I_{n_b} \\
&= \operatorname{tr}\{(L')^{m+1}\}I_{n_b} = n_d I_{n_b},
\end{aligned}
$$

since $\operatorname{tr}\{L^m\} = n_d$. We then find

$$
E[K_{13}] \overset{0}{=} 2n_d \operatorname{tr}\{E[M_+]\psi'\psi\}. \tag{61}
$$

Collecting terms we find from (59), (60), and (61) the expression for $E[\operatorname{tr}\{K_1\}]$.

K2: We have

$$
E[\operatorname{tr}\{K_2\}] = E\big[\operatorname{tr}\{M_{ab}(T^{-1}M_{bb\cdot d})^{-1}M_{by\cdot d}M_{yy\cdot d}^{-1}M_{yb\cdot d}(T^{-1}M_{bb\cdot d})^{-1}M_{ba}\}\big],
$$

and since $T^{-1}M_{bb\cdot d} \overset{0}{=} I_{n_b}$ and $M_{by\cdot d}M_{yy\cdot d}^{-1}M_{yb\cdot d} \overset{0}{=} M_{by}M_{yy}^{-1}M_{yb} \overset{0}{=} T^{-1}M_{by}M_{yb}$, we get since $M_{ba}M_{ab}$ is asymptotically independent of $M_{by}M_{yb}$ that from (48)

$$E[\operatorname{tr}\{K_2\}] \overset{0}{=} \operatorname{tr}\{E[M_{ba}M_{ab}]E[M_{by}M_{yb}]\} \overset{0}{=} n_y \operatorname{tr}\{E[M]\}.$$

K3: The term K_3 is given by

$$K_3 = M_{ab}(T^{-1}M_{bb.d})^{-1}M_{ba}M_{aa}^{-1}(M_{ay}(T^{-1}M_{yy.d})^{-1}M_{ya})M_{aa}^{-1}.$$

Again we can drop the conditioning on d_t and replace $T^{-1}M_{yy}$ by I_{n_y} and $T^{-1}M_{bb.d}$ by I_{n_b} and find

$$E[\operatorname{tr}\{K_3\}] \overset{0}{=} \operatorname{tr}\{E[M_{ab}M_{ba}M_{ay}M_{ya}]\}.$$

We condition on $B = \{B_t\}$ and use $M_{aa} = I_{n_a}$ and $\Sigma = I_{n_y}$ to find from (50)

$$\begin{aligned} E[\operatorname{tr}\{K_3\}] &\overset{0}{=} \operatorname{tr}\{E[M_{ab}M_{ba}(M_{ab}^{+}\psi'\psi M_{ba}^{+} + I_{n_a}\operatorname{tr}\{\theta'\theta\})]\} \\ &\overset{0}{=} \operatorname{tr}\{E[M_{+}'M_{+}]\psi'\psi\} + \operatorname{tr}\{E[M]\}v_\theta \end{aligned}$$

with $M = M_{ab}M_{ba}$, $M_{+} = M_{ba}M_{ab}^{+}$, and $v_\theta = \operatorname{tr}\{\theta'\theta\}$.

K4: The expression for K_4 is

$$T^{-1}M_{ay}(T^{-1}M_{yy.d})^{-1}M_{yb.d}(T^{-1}M_{bb.d})^{-1}M_{by.d}(T^{-1}M_{yy.d})^{-1}M_{ya}.$$

We drop the conditioning on d_t and replace $T^{-1}M_{yy}$ by I_{n_y} and $T^{-1}M_{bb.d}$ by I_{n_b}, and find from Lemma 3, by conditioning on $B = \{B_t\}$,

$$\begin{aligned} E[\operatorname{tr}\{K_4\}] &\overset{0}{=} T^{-1}\operatorname{tr}\{E[M_{ay}M_{yb}M_{by}M_{ya}]\} \\ &\overset{0}{=} \operatorname{tr}\{E[T^{-1}M_{yb}M_{by}]E[M_{ya}M_{ay}]\} \\ &\overset{0}{=} \operatorname{tr}\{n_b I_{n_y}(\psi M_{ba}^{+}M_{ab}^{+}\psi' + n_a\theta\theta')\} \\ &= n_b(n_a v_\theta + \operatorname{tr}\{E[M_{++}]\psi'\psi\}). \end{aligned}$$

This concludes the proof of Theorem 1. *Q.E.D.*

REFERENCES

ABADIR, K. M., K. HADRI, AND E. TZAVALIS (1999): "The Influence of VAR Dimension on Estimator Biases," *Econometrica*, 67, 163–181.

AHN, S. K., AND C. G. REINSEL (1990): "Estimation for Partially Nonstationary Multivariate Autoregressive Models," *Journal of the American Statistical Association*, 85, 813–823.

ANDERSON, T. W. (1951): "Estimating Linear Restrictions on Regression Coefficients for Multivariate Normal Distributions," *Annals of Mathematical Statistics*, 22, 327–351.

BARTLETT, M. S. (1937): "Properties of Sufficiency and Statistical Tests," *Proceedings of the Royal Society of London*, Series A, 160, 268–282.

BASU, D. (1955): "On Statistics Independent of a Complete Sufficient Statistic," *Sankhyā*, 15, 377–380.

BRAVO, F. (1999): "A Correction Factor for Unit Root Test Statistics," *Econometric Theory*, 15, 218–237.

CHEUNG, Y.-W., AND K. S. LAI (1993): "Finite Sample Sizes of Johansens Likelihood Ratio Tests for Cointegration," *Oxford Bulletin of Economics and Statistics*, 55, 313–332.

DOORNIK, J. A., B. NIELSEN, AND T. J. ROTHENBERG (2002): "The Influence of VAR Dimension on Estimator Biases: Comment," forthcoming *Econometrica*.

ENGLE, R. F., AND C. W. J. GRANGER (1987): "Co-integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55, 251–276.

GONZALO, J., AND J.-Y. PITARAKIS (1999): "Dimensionality Effect in Cointegration Analysis," in *Cointegration, Causality and Forecasting: Festschrift in Honour of Clive Granger*, ed. by R. F. Engle and H. White. Oxford: Oxford University Press, 212–229.

HANSEN, H., AND A. C. RAHBEK (2002): "Approximate Conditional Unit Root Inference," *Journal of Times Series Analysis*, 23, 1–18.

HAUG, A. A. (1996): "Tests for Cointegration: A Monte Carlo Comparison," *Journal of Econometrics*, 71, 89–115.

JACOBSON, T., AND R. LARSSON (1999): "Bartlett Correction in Cointegration Testing," *Computational Statistics and Data Analysis*, 31, 203–225.

JENSEN, J. L., AND A. T. A. WOOD (1997): "On the Non-existence of a Bartlett Correction for Unit Root Tests," *Statistics and Probability Letters*, 35, 181–187.

JOHANSEN, S. (1988): "Statistical Analysis of Cointegration Vectors," *Journal of Economic Dynamics and Control*, 12, 231–254.

——— (1991): "The Power Function of the Likelihood Ratio Test for Cointegration," in *Econometric Decision Models: New Methods of Modeling and Applications*, ed. by J. Gruber. New York: Springer Verlag, 323–335.

——— (1996): *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford University Press.

——— (2000): "A Bartlett Correction Factor for Tests on the Cointegrating Relations," *Econometric Theory*, 16, 740–778.

——— (2002): "A Small Sample Correction for Tests of Hypotheses on the Cointegrating Vectors," forthcoming in *Journal of Econometrics*.

JOHANSEN, S., H. HANSEN, AND S. FACHIN (2002): "A Simulation Study of Some Functionals of Random Walk," Discussion Paper, www.math.ku.dk/~sjo.

LARSSON, R. (1997): "On the Asymptotic Expectations of Some Unit Root Tests in a First Order Autoregressive Process in the Presence of Trend," *Annals of the Institute of Statistical Mathematics*, 49, 585–599.

——— (1998): "Bartlett Corrections for Unit Root Test Statistics," *Journal of Times Series Analysis*, 19, 426–438.

LAWLEY, D. N. (1956): "A General Method for Approximating the Distribution of Likelihood Ratio Criteria," *Biometrika*, 43, 296–303.

NIELSEN, B. (1997a): "Bartlett Correction of the Unit Root Test in Autoregressive Models," *Biometrika*, 84, 500–504.

——— (1997b): "On the Distribution of Test of Cointegration," 1997-W10, Nuffield College, Oxford, www.nuff.ox.ac.uk/users/nielsen.

REIMERS, H.-E. (1992): "Comparison of Tests for Multivariate Cointegration," *Statistical Papers*, 33, 335–359.

TODA, H. Y. (1995): "Finite Sample Performance of Likelihood Ratio Tests for Cointegrating Rank in Vector Autoregressions," *Econometric Theory*, 11, 1015–1032.