A Small Sample Correction of the Test for Cointegrating Rank in the Vector Autoregressive Model

Søren Johansen

A small sample correction of the test for cointegrating rank in the vector autoregressive model

Søren Johansen,
Economics Department,
European University Institute, Florence

May 2000

Abstract

We derive an approximation to the expectation of the likelihood ratio test for cointegration in the vector autoregressive model. The expression depends on moments of functions of 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 the usefulness of the correction is illustrated by some simulations experiments.

1 Introduction and models

In the vector autoregressive model the likelihood ratio test for cointegration rank is the so called trace test derived by the technique of reduced rank regression, Anderson (1951). The asymptotic distribution under the assumption of $I(1)$ is derived by Johansen (1988, 1996) and Ahn and Reinsel (1990). There are many studies that show that for small samples the size of the test is not well approximated by the asymptotic value, see for example Cheung and Lai (1993), Toda (1995), Gonzalo and Pitarakis (1999), and Haug (1996). Reimers (1992) proposed a small sample correction based on degrees of freedom, and Hansen and Rahbek (2000) employ ideas of profile likelihood to derive corrections of the Dickey Fuller test.

In this paper we propose a correction factor to the likelihood ratio test which improves the finite sample properties. The idea is that of the Bartlett correction, see Bartlett (1937). Bartlett suggested finding an expression for the expectation of the likelihood ratio test statistic. By means of this the statistic is corrected to have the same mean as the limit distribution, thereby improving the approximation.

A more precise formulation is as follows. We let θ denote the parameter and want to test the hypothesis of cointegration by the likelihood ratio test, LR. We find an expression of the form

$$E_\theta(-2\log LR) = E(T)(1 + T^{-1}b(\theta) + \cdots),$$

where $E(T)$ is the expectation of the LR for no cointegration in a model with one lag. In the models considered in this paper the expectation, $E(T)$, depends on the sample size, the number of common trends, and the type of deterministic terms in the model, but not on the parameters. Let $E(\infty) = \lim_{T\to\infty} E(T)$ be the mean of the limit distribution. The correction takes the form

$$\frac{E(\infty)}{E(T)}\frac{-2\log LR}{1 + T^{-1}b(\hat\theta)} = \frac{-2\log LR}{a(T)(1 + T^{-1}b(\hat\theta))},$$

which, at least to the order T^{-1}, has the same mean as the limit distribution, and hence presumably has better finite sample properties. Here $a(T) = E(T)/E(\infty)$. In this paper we derive an analytic expression for $b(\theta)$ and derive by simulation an expression for $a(T)$.

The distribution of the likelihood ratio test statistic depends on sample size, T, and parameter, θ. For $T \to \infty$ the limit is not approached uniformly

in the parameter, since the convergence is slow if θ is close to a boundary point where the cointegrating properties change. This non-uniformity in the parameter causes the approximation to be poor in certain areas of the parameter space, see Nielsen (1997b) for a discussion. In some examples where we can calculate the correction term we can quantify this phenomenon.

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), but in the unit root case we should not expect as much, see Jensen and Wood (1995) and Bravo (1998), even though Nielsen (1997a) shows that one in practice can get a better fit.

The rest of this section defines the models and the next section gives the main result for the approximation of the expectation of the likelihood ratio test and the proposed correction factor. Section 3 contains some simple models where the correction can be worked out explicitly and the results are illustrated by simulation. In Section 4 we discuss the calculation of the coefficients needed for the correction factor. In Section 5 the main steps of the proof are given, but most details are referred to an Appendix.

1.1   The models

We work throughout with the autoregressive model $\mathcal{M}_1$ for the $n-$dimensional process X_t, $t = 1, \ldots, T$ defined by

$$\mathcal{M}_1 : \Delta X_t = \Pi X_{t-1} + \Upsilon D_t + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi d_t + \varepsilon_t, \qquad (1)$$

where the errors ε_t are i.i.d. $N_n(0, \Omega)$ and the initial conditions are fixed. We consider the deterministic terms $D_t = t^{n_d}$ and $\Phi d_t = \sum_{i=0}^{n_d - 1} \Phi_i t^i$. The parameters Π, $\Gamma_1, \ldots, \Gamma_{k-1}$, are $(n \times n)$, Φ is $(n \times n_d)$, and Υ $(n \times 1)$. The null hypothesis we want to test is

$$\Pi : \Pi = \alpha \beta', \Upsilon = \alpha \rho'.$$

Under the null hypothesis the model is

$$\mathcal{M}_2 : \Delta X_t = \alpha(\beta' X_{t-1} + \rho' t^{n_d}) + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi d_t + \varepsilon_t, \qquad (2)$$

where all parameters are unrestricted, but now 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. These models correspond to the trend t^{n_d} being restricted to the cointegration space. We formulate the main result in Theorem 1 about $E[-2 \log LR(\mathcal{M}_2|\mathcal{M}_1)]$ to cover the case of $D_t = 0$, (or $\Upsilon = 0$) such that we only test $\Pi = \alpha\beta'$. We will implement the correction only in the case $D_t = t^{n_d}$.

Of particular interest is the hypothesis $\Pi = 0$ and $\Upsilon = 0$, in the model with only one lag

$$\Delta X_t = \Pi X_{t-1} + \Upsilon t^{n_d} + \Phi d_t + \varepsilon_t,$$

which gives the Dickey Fuller test in n dimensions. The expected likelihood ratio test statistic is

$$E(T, n, n_d) = -2TE[\log|I_n - M_{\varepsilon\varepsilon.d}^{-1} M_{\varepsilon\tilde{x}.d} M_{\tilde{x}\tilde{x}.d}^{-1} M_{\tilde{x}\varepsilon.d}|], \tag{3}$$

where we have used the product moment notation

$$M_{\varepsilon\tilde{x}.d} = \sum_{t=1}^{T} \varepsilon_t \left(\left. \begin{array}{c} X_{t-1} \\ t^{n_d} \end{array} \right| d_t \right)' = \sum_{t=1}^{T} \varepsilon_t \left(\left. \begin{array}{c} X_0 + \sum_{i=1}^{t-1}(\varepsilon_i + \Phi d_i) \\ t^{n_d} \end{array} \right| d_t \right)',$$

and where $(U_t|V_t)$ denotes the residuals of a regression of U_t on V_t. The distribution in (3) does not involve any parameters but only sample size T and dimensions n and n_d. The exact analytic expression for $E(T, n, n_d)$ is difficult, see Larsson (1998a,1998b), Nielsen (1997a), and Abadir, Hadri and Tzavalis (1999), and we propose to tabulate it as a function of T, n, and n_d, and apply it as normalization in the correction factor.

2   The main result

In order to formulate the results we need a notation for some functions of the parameters and some product moment matrices derived from a random walk.

2.1   Parameter functions

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

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

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

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

with

$$P = \begin{pmatrix} I_r + \beta'\alpha & \beta'\Gamma_1 & \cdots & \beta'\Gamma_{k-2} & \beta'\Gamma_{k-1} \\ \alpha & \Gamma_1 & \cdots & \Gamma_{k-2} & \Gamma_{k-1} \\ 0 & I_n & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & I_n & 0 \end{pmatrix}, Q = \begin{pmatrix} \beta' \\ I_n \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad (4)$$

and variance $\Sigma = \mathrm{Var}(Y_t)$.

We define the permanent shocks

$$B_t = (\alpha_\perp'\Omega\alpha_\perp)^{-\frac{1}{2}}\alpha_\perp'\varepsilon_t \qquad (5)$$

of dimension $n_b = n - r$, and the transitory shocks

$$U_t = (\alpha'\Omega^{-1}\alpha)^{-\frac{1}{2}}\alpha'\Omega^{-1}\varepsilon_t \qquad (6)$$

of dimension $n_u = r$, such that $(B_t', U_t')'$ are i.i.d. $N_n(0, I_n)$, and find the representation

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

with

$$\theta_i = P^i Q\alpha(\alpha'\Omega^{-1}\alpha)^{-\frac{1}{2}}, \; \psi_i = P^i Q\Omega\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-\frac{1}{2}}.$$

We find the variances and covariances

$$\begin{aligned}
&\Sigma_\theta = \mathrm{Var}(Y_{\theta t}) = \sum_{i=0}^{\infty}\theta_i\theta_i', \; \Sigma_\psi = \mathrm{Var}(Y_{\psi t}) = \sum_{i=0}^{\infty}\psi_i\psi_i', \\
&\Sigma = \mathrm{Var}(Y_t) = \sum_{i=0}^{\infty}(\theta_i\theta_i' + \psi_i\psi_i') = \Sigma_\theta + \Sigma_\psi, \\
&\gamma_\theta(h) = \mathrm{Cov}(Y_{\theta t}, Y_{\theta.t+h}) = \Sigma_\theta P'^h, \gamma_\psi(h) = \mathrm{Cov}(Y_{\psi t}, Y_{\psi.t+h}) = \Sigma_\psi P'^h, \\
&\gamma(h) = \mathrm{Cov}(Y_t, Y_{t+h}) = \Sigma P'^h = \gamma_\psi(h) + \gamma_\theta(h).
\end{aligned} \qquad (7)$$

We define the long-run coefficients

$$\begin{aligned}
\theta &= \sum_{i=0}^{\infty}\theta_i = (I_{n_y} - P)^{-1}Q\alpha(\alpha'\Omega^{-1}\alpha)^{-\frac{1}{2}}, \\
\psi &= \sum_{i=0}^{\infty}\psi_i = (I_{n_y} - P)^{-1}Q\Omega\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-\frac{1}{2}},
\end{aligned}$$

4

and find the long-run variances $\theta\theta'$ and $\psi\psi'$ of $Y_{\theta t}$ and $Y_{\psi t}$ respectively. Let

$$\begin{aligned}
V_\theta &= \theta\theta'\Sigma^{-1} = (I_{n_y} - P)^{-1}Q\alpha(\alpha'\Omega^{-1}\alpha)^{-1}\alpha'Q'(I_{n_y} - P')^{-1}\Sigma^{-1}, \\
V_\psi &= \psi\psi'\Sigma^{-1} = (I_{n_y} - P)^{-1}Q\Omega\alpha_\perp(\alpha'_\perp\Omega\alpha_\perp)^{-1}\alpha'_\perp\Omega Q'(I_{n_y} - P')^{-1}\Sigma^{-1}.
\end{aligned} \tag{8}$$

Finally we need the matrix

$$V = \psi'\Sigma^{-1}\sum_{h=0}^\infty \psi_h tr\{\Sigma^{-1}\gamma(h+1)\} + \psi'\Sigma^{-1}\sum_{h=0}^\infty \gamma(h+1)'\Sigma^{-1}\psi_h. \tag{9}$$

2.2 Product moments of random walks

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

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

and for the case $D_t = 0$ we define $A_{t-1} = \left(\sum_{i=1}^{t-1} B_i + \mu t^{n_d} \middle| d_t \right)$, of dimension $n_a = n - r$, where $\mu = n_d^{-1}(\alpha'_\perp\Omega\alpha_\perp)^{-\frac{1}{2}}\alpha'_\perp\Phi_{n_d-1}$ is the coefficient of the trend t^{n_d} in $(\alpha'_\perp\Omega\alpha_\perp)^{-\frac{1}{2}}\alpha'_\perp X_t$. We define the product moments

$$\begin{aligned}
M_{aa} &= \sum_{i=1}^T A_{t-1}A'_{t-1}, \; M_{ab} = \sum_{i=1}^T A_t B'_{t-1}, \; M_{ab}^+ = \sum_{i=1}^T A_t B'_t, \\
M_{bb} &= \sum_{i=1}^T B_t B'_t, \; M_{bb.d} = M_{bb} - M_{bd}M_{dd}^{-1}M_{db},
\end{aligned} \tag{10}$$

and

$$\begin{aligned}
M_1 &= M_{ba}M_{aa}^{-1}M_{ab}, \\
M_2 &= (M_{ba}M_{aa}^{-1}M_{ab})^2 = M_1^2, \\
M_3 &= M_{ba}M_{aa}^{-1}M_{ab}^+, \\
M_4 &= M_{ba}^+M_{aa}^{-1}M_{ab}M_{ba}M_{aa}^{-1}M_{ab}^+ = M_3'M_3, \\
M_5 &= M_{bb}^-M_{ba}M_{aa}^{-1}M_{ab}^+, \\
M_6 &= M_{ba}^+M_{aa}^{-1}M_{ab}^+.
\end{aligned} \tag{11}$$

We then formulate the main result about an approximation of the expectation of the log likelihood ratio test.

Theorem 1 *The expectation of the test for $\Pi = \alpha\beta', \Upsilon = \alpha\rho'$ when $D_t = t^{n_d}$, (or $\Pi = \alpha\beta'$ if $D_t = 0$) in model (1) has the expansion*

$$\begin{aligned}
E[-2logLR(\mathcal{M}_2|\mathcal{M}_1)] &\stackrel{1}{=} -2TE[\log|I_{n_b} - M_{bb.d}^{-1}M_{ba}M_{aa}^{-1}M_{ab}|] \\
&+ T^{-1}(tr\{M_1\}tr\{\psi'\Sigma^{-1}\psi\} + tr\{M_3\}tr\{I_{n_y} - \theta\theta'\Sigma^{-1} - \psi\psi'\Sigma^{-1}\}) \\
&+ 2T^{-1}(tr\{E[M_3]V\} - n_d tr\{E[M_3]\psi'\Sigma^{-1}\psi\}) \\
&+ T^{-1}tr\{E[M_4 - 2M_5 + (n - r)M_6]\psi'\Sigma^{-1}\psi\}.
\end{aligned}$$

Here the moments M_i and M_{ij} are given in (10) and (11) and the coefficients θ, ψ, Σ, V are found in (7), (8), and (9).

5

The proof will be given in the Appendix. We have used the notation $\overset{1}{=}$ to indicate that we have kept terms of order T^{-1}.

We next apply this to the situation $D_t = t^{n_d}$, where the first term does not depend on any parameters and we define the coefficients $E(T, n_b, n_d)$, $a(T, n_b, n_d)$, and $\gamma(n_b, n_d)$ by

$$
\begin{aligned}
E(T, n_b, n_d) &= -2TE[\log|I_{n_b} - M_{ba}M_{aa}^{-1}M_{ab}M_{bb.d}^{-1}|] \\
a(T, n_b, n_d) &= E(T, n_b, n_d)/E(\infty, n_b, n_d) \\
\gamma(n_b, n_d) &= n_b \lim_{T \to \infty} E[tr\{M_{ba}M_{aa}^{-1}M_{ab}^+\}]/E[tr\{M_{ba}M_{aa}^{-1}M_{ab}\}]
\end{aligned}
\tag{12}
$$

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

Corollary 2 *The correction factor for the test of $\mathcal{M}_2$ in $\mathcal{M}_1$, that is, the test for cointegrating rank r in the vector autoregressive model (1) with $D_t = t^{n_d}$ and $d_t = 1, \ldots, t^{n_d-1}$, $n_d = 0, 1, 2$ is*

$$
a(T, n - r, n_d)(1 + \frac{1}{T}[c_1 + \{(n - r)c_2 + 2(c_3 - n_dc_1)\}\frac{\gamma(n - r)}{(n - r)^2}]).
$$

The coefficient $a(T, n - r, n_d)$ is well approximated by

$$
a(T, n - r, n_d) = 1 + 0.45T^{-1}(n - r) + 1.2T^{-2}(n - r)^2
$$

and approximations of $\gamma(n - r, n_d)$ are given in Table 1 for $n_d = 0, 1, 2$. Finally

$$
\begin{aligned}
c_1 &= tr\{V_\psi\} = tr\{(I_{n_y} - P)^{-1}Q\Omega\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}\alpha_\perp'\Omega Q'(I_{n_y} - P')^{-1}\Sigma^{-1}\}, \\
c_2 &= tr\{I_{n_y} - (I_{n_y} - P)^{-1}Q\Omega Q'(I_{n_y} - P')^{-1}\Sigma^{-1}\}, \\
c_3 &= tr\{V\} = tr\{[(I_{n_y} - P)V_\psi \otimes P][I_{n_y^2} - P \otimes P]^{-1}\} + tr\{V_\psi P(I_{n_y} + P)^{-1}\}.
\end{aligned}
$$

The functions of the parameters are described in Section 2.1.

Proof. We find from (32) and (33) of Lemma 4 in the Appendix, that

$$
\begin{aligned}
\lim_{T \to \infty} E(T, n_b, n_d) &= m_1(n_b, n_d), \\
\lim_{T \to \infty} E[tr\{M_3\psi'\Sigma^{-1}\psi\}] &= m_3(n_b, n_d)n_b^{-1}c_1, \\
\lim_{T \to \infty} E[tr\{M_3V\}] &= m_3(n_b, n_d)n_b^{-1}c_3, \\
\lim_{T \to \infty} E[M_4 - 2M_5 + (n - r)M_6] &= 0, \\
\gamma(n_b, n_d) &= n_b m_3(n_b, n_d)/m_1(n_b, n_d),
\end{aligned}
$$

6

	n_d	d_t	D_t	$a(T,n,n_d)$	1	n/T	$(n/T)^2$
					1.000	0.450	1.200
		0	0	$\gamma(n)$	-0.499	0.069	-0.120
	0	0	1	$\gamma(n,0)$	-0.486	0.180	-0.016
	1	1	t	$\gamma(n,1)$	-1.469	1.457	-0.611
	2	$1,t$	t^2	$\gamma(n,2)$	-2.338	3.306	-1.702

Table 1: Simple approximations of $a(T,n_b,n_d)$ and $\gamma(n,n_d)$ for the model with $D_t = t^{n_d}$. The approximations are found by fitting a curve to simulated values based upon 100.000 simulations and values of $n = 1,\dots,10$, $n_d = 0,1,2$, and $T = 50,100,150,200,300,400,500,1000$. We approximate $a(T,n_b,n_d)$ by $E(T,n_b,n_d)/E(1000,n_b,n_d)$, and define $\gamma(n)$ by (12) for the model with no deterministic terms.

and hence with $n_b = n - r$

$$E[-2\log LR(\mathcal{M}_2|\mathcal{M}_1)]$$
$$\overset{1}{=} E(T,n_b,n_d) + T^{-1}(m_1(n_b,n_d)c_1 + m_3(n_b,n_d)c_2 + 2m_3(n_b,n_d)n_b^{-1}(c_3 - n_dc_1)$$
$$\overset{1}{=} E(T,n_b,n_d)(1 + T^{-1}(c_1 + \gamma(n_b,n_d)n_b^{-2}(n_bc_2 + 2(c_3 - n_dc_1)))$$
$$= E(\infty,n_b,n_d)a(T,n_b,n_d)(1 + T^{-1}(c_1 + \gamma(n_b,n_d)n_b^{-2}(n_bc_2 + 2(c_3 - n_dc_1))),$$

which is the result given in Corollary 2. ∎

The matrix appearing in the coefficient c_2 is 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 n_d, but on the parameters and hence the dimensions n and r.

3 Some special cases

We next illustrate the results in some special cases, where the coefficients can be worked out explicitly and which are convenient for simulation experiments and 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 with $k = 2, n_d = 1$

$$\Delta X_t = \Pi X_{t-1} + \Gamma_1 \Delta X_{t-1} + \Upsilon t + \Phi + \varepsilon_t \tag{13}$$

where ε_t are i.i.d. $N_n(0, \Omega)$, Π and Γ_1 are $n \times n$, Υ and Φ are $n \times 1$. Under the null hypothesis there is no cointegration, but parameters Γ_1 and Φ, which generates a linear trend in the process. We can find a simple expression for the correction factor, if we assume that $\Gamma_1 = \xi I_n$, and hence see the effect of the short term dynamics. In this case we have $\alpha = \beta = 0$, $\alpha_\perp = \beta_\perp = I_n$, and that $Y_t = X_t - E(\Delta X_t)$ is autoregressive with coefficient $P = \xi I_n$ and $Q = I_n$, and hence has the representation

$$Y_t = \sum_{i=0}^{\infty} \xi^i \varepsilon_{t-i}.$$

This gives

$$\Sigma = \text{Var}(\Delta X_t) = \sum_{i=0}^{\infty} \xi^{2i} \Omega = \frac{1}{1-\xi^2} \Omega, \; V_\psi = \frac{1+\xi}{1-\xi} I_n,$$

such that

$$c_1 = n\frac{1+\xi}{1-\xi}, \; c_2 = -2n\frac{\xi}{1-\xi}, \; c_3 = n(n+1)\frac{\xi}{1-\xi}.$$

From Corollary 2 we get the correction factor with $n_y = n, n_d = 1, k = 2, r = 0$

$$a(T, n, 1)(1 + \frac{1}{T}[n\frac{(1+\xi)}{1-\xi} - \frac{2\gamma(n,1)}{n(1-\xi)}]). \tag{14}$$

If instead $d_t = 0$ and $D_t = 1$, that is, $n_d = 0$, we find the correction factor

$$a(T, n, 0)(1 + \frac{1}{T}[n\frac{1+\xi}{1-\xi} + \frac{2\xi\gamma(n,0)}{n(1-\xi)}]). \tag{15}$$

Some simulations were performed and are given in Table 2 to illustrate the usefulness of formula (14). The DGP has $n = 5$, $\Pi = 0, \Upsilon = 0, \Gamma_1 = \xi I_n, \Omega = I_n$. Note that as ξ tends to 1 the size of the test increases to 1 and hence the asymptotic tables for the trace test are not useful. The correction factor manages to correct the size to a reasonable level for $\xi \leq 0.6$, say. The limit for ξ tending to 1 corresponds to the process being $I(2)$. For $\xi = 0.6$, a nominal 5% test, using the asymptotic critical values is in reality a 84% test. The correction brings the size down to 2%. As the size becomes even more distorted the correction factor overcorrects due to the singularity in the expression $1/(1-\xi)$.

Thus the parametric expression for the correction allows us to discuss when the approximation using the asymptotics is useful and when the correction is needed, and finally when both are useless.

$T\backslash\xi$	0	0.3	0.5	0.6	0.7
50	$\frac{37.8}{7.8}(1.17)$	$\frac{55.8}{5.7}(1.27)$	$\frac{73.6}{3.7}(1.39)$	$\frac{84.2}{2.0}(1.51)$	$\frac{91.9}{0.004}(1.69)$
100	$\frac{17.5}{6.9}(1.08)$	$\frac{25.1}{6.8}(1.13)$	$\frac{35.0}{4.7}(1.19)$	$\frac{44.0}{4.2}(1.24)$	$\frac{58.7}{3.0}(1.33)$

Table 2: In model (13) with $T = 50, n = 5$, we simulate the test that $\Pi = 0, \Upsilon = 0$. The number of simulations is 10.000. The correction is calculated from (14). The table shows the simulated size of a nominal 5% test using asymptotic critical values over the corrected size, with the factor in parenthesis.

3.2 The test for rank one in the model with one lag

The model with $k = 1, r = 1, n_d = 1$, is

$$\Delta X_t = \Pi X_{t-1} + \Upsilon t + \Phi + \varepsilon_t, \tag{16}$$

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

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

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

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

and find the coefficients

$$c_1 = -\frac{(2+\beta'\alpha)}{\beta'\alpha}\kappa, \ c_2 = \frac{2(1+\beta'\alpha)}{\beta'\alpha}, c_3 = -2\frac{(1+\beta'\alpha)}{\beta'\alpha}\kappa.$$

From Corollary 2 we find the correction factor for testing $r = 1$ in the model with only one lag:

$$a(T, n-1, 1)$$
$$\times(1 + \frac{1}{T}[\frac{-(2+\beta'\alpha)}{\beta'\alpha}\kappa + \{\frac{2(1+\beta'\alpha)}{\beta'\alpha}(n-1) - 2\kappa\}\frac{\gamma(n-1,1)}{(n-1)^2}]).$$

If instead we take $d_t = 0, D_t = 1$, and $n_d = 0$, we get

$$a(T, n-1, 0)$$
$$\times(1 + \frac{1}{T}[-\frac{(2+\beta'\alpha)}{\beta'\alpha}\kappa + \frac{2(1+\beta'\alpha)}{\beta'\alpha}\{(n-1) - 2\kappa\}\frac{\gamma(n-1,0)}{(n-1)^2}]).$$

We see that the formula breaks down if $\alpha'\beta = 0$, and that corresponds again to X_t being $I(2)$, see Johansen (1996, Theorem 4.2).

9

$T = 100$				
$a_2\backslash a_1$	-0.2	-0.4	-0.6	-0.8
-0.2	$\frac{0.06}{0.01}$ (1.10)	$\frac{0.07}{0.04}$ (1.04)	$\frac{0.09}{0.06}$ (1.03)	$\frac{0.08}{0.06}$ (1.03)
-0.4	$\frac{0.12}{0.03}$ (1.13)	$\frac{0.10}{0.05}$ (1.06)	$\frac{0.09}{0.06}$ (1.04)	$\frac{0.09}{0.07}$ (1.03)
-0.6	$\frac{0.15}{0.03}$ (1.14)	$\frac{0.11}{0.05}$ (1.06)	$\frac{0.09}{0.06}$ (1.04)	$\frac{0.09}{0.06}$ (1.03)
-0.8	$\frac{0.15}{0.04}$ (1.14)	$\frac{0.11}{0.05}$ (1.07)	$\frac{0.10}{0.06}$ (1.05)	$\frac{0.10}{0.07}$ (1.03)

Table 3: Test for $\Pi = \alpha\beta'$, $\Upsilon = \alpha\rho$ in model (16) with $T = 100, n = 5, r = 1$. The entries show the simulated size of a nominal 5% test over the corrected size and the factor in parenthesis. Number of simulations is 10.000 and the DGP has $\alpha = (a_1, a_2, 0, 0, 0)$, and $\beta = (1, 0, 0, 0, 0)$, and $\Omega = I_5$.

$T = 50$				
$a_2\backslash a_1$	-0.2	-0.4	-0.6	-0.8
-0.2	$\frac{0.05}{0.00}$ (1.21)	$\frac{0.05}{0.01}$ (1.09)	$\frac{0.08}{0.03}$ (1.07)	$\frac{0.10}{0.05}$ (1.06)
-0.4	$\frac{0.14}{0.00}$ (1.27)	$\frac{0.09}{0.02}$ (1.12)	$\frac{0.10}{0.03}$ (1.08)	$\frac{0.11}{0.05}$ (1.07)
-0.6	$\frac{0.21}{0.01}$ (1.29)	$\frac{0.14}{0.03}$ (1.14)	$\frac{0.12}{0.04}$ (1.09)	$\frac{0.12}{0.06}$ (1.07)
-0.8	$\frac{0.24}{0.01}$ (1.29)	$\frac{0.17}{0.04}$ (1.15)	$\frac{0.15}{0.05}$ (1.10)	$\frac{0.13}{0.07}$ (1.08)

Table 4: Test for $\Pi = \alpha\beta'$, $\Upsilon = \alpha\rho$ in model (16) with $T = 50, n = 5, r = 1$. The entries show the simulated size of a nominal 5% test over the corrected size and the factor in parenthesis. Number of simulations is 10.000 and the DGP has $\alpha = (a_1, a_2, 0, 0, 0)$, and $\beta = (1, 0, 0, 0, 0)$, and $\Omega = I_5$.

It is easy to see that in model (16) it is enough to consider simulation experiments with $\Omega = I_n, \beta = (1, 0, \dots, 0)$ and $a = (a_1, a_2, 0, \dots, 0)$. Thus Tables 3 and 4 suitably extended cover all possible simulation results for $n = 5, r = 1$. It is seen that for $a_1 + a_2 \leq -0.4$, the formula works reasonably well for $T = 100$, but for $T = 50$, we need $a_1 + a_2 \leq -1.0$.

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 model

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Upsilon t^{n_d} + \Phi d_t + \varepsilon_t. \tag{17}$$

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

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

We find

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

where e_i is the i-th 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 $E_{k-1}^{k-1} = (E_{k-1})^{k-1} = 0$, and that

$$P^i Q = (I_n \otimes E_{k-1}^i)(I_n \otimes e_1) = (I_n \otimes e_{i+1}), i = 0, \dots, k-1$$
$$(I_{n_y} - P)^{-1} Q = I_n \otimes (1, \dots, 1)' = I_n \otimes \iota,$$

11

say. We find for $\alpha_\perp = I_n$ that

$$V_\psi = (I_{n_y} - P)^{-1} Q \Omega \alpha_\perp (\alpha'_\perp \Omega \alpha_\perp)^{-1} \alpha'_\perp \Omega Q' (I_{n_y} - P')^{-1} \Sigma^{-1}$$
$$= [I_n \otimes \iota] \Omega [I_n \otimes \iota'][\Omega^{-1} \otimes I_{k-1}] = [I_n \otimes \iota\iota'],$$

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

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

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

$$c_3 = tr\{V_\psi P(I_{n_y} + P)^{-1}\} = \sum_{i=0}^\infty tr\{[I_n \otimes \iota\iota'][I_n \otimes E][I_n \otimes (-1)^i E^i]\}$$
$$= \sum_{i=0}^\infty (-1)^i tr\{[I_n \otimes \iota\iota' E^{i+1}]\} = \sum_{i=0}^\infty (-1)^i n \iota' E^{i+1} \iota$$
$$= n \sum_{i=0}^{k-2} (-1)^i (k-2-i) = n[\frac{k-1}{2}],$$

that is, n times the integer part of $(k-1)/2$. Thus

$$c_1 = (k-1)n, \ c_2 = 0, \ c_3 = n[\frac{k-1}{2}].$$

From Corollary 2 we get the result

$$a(T, n, n_d)(1 + \frac{1}{T}[(k-1)n + \{[\frac{k-1}{2}] - n_d(k-1)\}\frac{2\gamma(n, n_d)}{n}]). \qquad (18)$$

This coincides with (14) and (15) for $k = 2$ and $\xi = 0, n_d = 0, 1$. For $k = 1$ we just get the correction factor one $a(T, n, n_d)$, corresponding to the normalization on the Dickey Fuller test in the model with one lag. Further simulations, see for instance Table 5, show that as long as the number of parameters per observation, kn/T, is less than 0.2, the formula gives a good approximation. Thus for instance for $n = 5, k = 2, T = 50$ a nominal 5% test using asymptotic critical values is in fact a 38.8% test. The correction gives a test with size 7.9%, which is close enough to the 5% we want, and much better than the direct use of the asymptotic tables.

Note that the first term of the correction $(1 + \frac{1}{T}(k-1)n)$ corresponds to multiplying the likelihood ratio test by $T^{-1}(T - (k-1)n)$, which is the correction found by Hansen and Rahbek (2000) based on a profile likelihood, whereas Reimers (1992) suggested to use $T^{-1}(T - kn)$.

$k\backslash T$	50	100	500
1	$\frac{14.4}{7.4}(1.05)$	$\frac{9.7}{7.0}(1.02)$	$\frac{7.2}{6.8}(1.00)$
2	$\frac{38.8}{7.9}(1.17)$	$\frac{17.3}{6.8}(1.08)$	$\frac{8.6}{6.9}(1.02)$
3	$\frac{70.5}{12.9}(1.28)$	$\frac{28.9}{7.7}(1.13)$	$\frac{8.7}{6.4}(1.03)$
4	$\frac{93.2}{24.2}(1.40)$	$\frac{43.6}{8.3}(1.19)$	$\frac{10.8}{6.8}(1.04)$

Table 5: Test for $\Pi = 0, \Upsilon = 0$ in model (17), when $n = 5$ and $\Gamma_i = 0$. Entries are simulated size over corrected size of a nominal 5% test and the factor in parenthesis. The correction is calculated using (18). The number of simulations is 10.000.

3.4   A real life examples

We illustrate the methods by a data sets 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 a 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 size of the tests which use the asymptotic distribution, and the effect of applying the correction factor.

In each simulation we use the estimated values of the parameters 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 leave out the parameters $\hat{\alpha}$ and $\hat{\beta}$ 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 $\hat{\beta}$ from the Danish data.

We compare in Table 6 the simulated 95% quantiles with the asymptotic ones. The simulated are larger than the asymptotic ones and this is what

r	n-r	95%(asym)	95%(sim)	
0	4	53.42	61.85	$\frac{19.2}{6.3}(1.11)$
1	3	34.80	40.79	$\frac{17.5}{4.3}(1.27)$

Table 6: The correction of the rank test for the Danish data. The columns give the rank and common trends tested, the asymptotic 95% quantiles and the simulated ones. Next the simulated size over the corrected size of a nominal 5% test and the factor in parenthesis.

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% test instead of the nominal 5% size. The corrected test has a size of 6% test. For $r = 1$ we get a 17% test instead of 5%, but the correction factor brings the size down to 4%.

When we simulate the DGP with $r = 1$, and use the test statistic for $r = 0$, we find that a 5% test has power 69% and a 10% test has power 81%.

4 Calculation of coefficients in Theorem 1

We have given in Section 3 some examples where the correction has a simple expression. We next discuss the general expressions and their calculation.

4.1 Calculation of variance

The variance Σ can be found from the linear equations

$$\Sigma = P\Sigma P' + Q\Omega Q',$$

with solution

$$vec(\Sigma) = (I_{n_y^2} - P \otimes P)^{-1}vec(Q\Omega Q').$$

The matrix to be inverted above can be quite large $(n_y^2 \times n_y^2)$, and it is sometimes an advantage to diagonalize $P(n_y \times n_y)$ and use a different expression for the variance. Let $P = KRK^{-1}$ where $R = diag(\rho_1, \dots, \rho_{n_y})$ so that

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

14

and hence

$$\Sigma = K\{(K^{-1}Q\Omega Q'K'^{-1})_{ij}/(1 - \rho_i\rho_j)\}K'.$$

For many simple examples the matrix P cannot be diagonalized but need a Jordan decomposition. For the case where the parameters are estimated, however, this happens with probability zero. In general P will have complex eigenvalues and eigenvectors.

4.2   The coefficient c_1, c_2, and c_3

The coefficients c_1 and c_2 are calculated directly from their expressions in Corollary 2. The coefficient $c_3 = tr\{V\}$, see (9), is given by

$$c_3 = tr\{\psi'\Sigma^{-1}\sum_{h=0}^{\infty}\psi_h tr\{\Sigma^{-1}\gamma(h+1)\} + \psi'\Sigma^{-1}\sum_{h=0}^{\infty}\gamma(h+1)'\Sigma^{-1}\psi_h\}. \quad (19)$$

We first evaluate

$$tr\{\psi'\Sigma^{-1}\sum_{h=0}^{\infty}\psi_h\}tr\{\Sigma^{-1}\gamma(h+1)\}$$
$$= \sum_{h=0}^{\infty} tr\{Q\Omega\alpha_{\perp}(\alpha_{\perp}'\Omega\alpha_{\perp})^{-1}\alpha_{\perp}'\Omega_1 Q'(I_{n_y} - P')^{-1}\Sigma^{-1}P^h\}tr\{P^{h+1}\}$$
$$= \sum_{h=0}^{\infty} tr\{(I - P)V_{\psi}P^h\}tr\{P^{h+1}\}$$
$$= \sum_{h=0}^{\infty} tr\{[(I_{n_y} - P)V_{\psi} \otimes P][P^h \otimes P^h]\}$$
$$= tr\{[(I_{n_y} - P)V_{\psi} \otimes P][I_{n_y^2} - P \otimes P]^{-1}\}.$$

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

$$\sum_{h=0}^{\infty} tr\{(I - P)V_{\psi}P^h\}tr\{P^{h+1}\}$$
$$= \sum_{h=0}^{\infty} tr\{K(I - R)K^{-1}V_{\psi}KR^h K^{-1}\}tr\{R^{h+1}\} = \sum_{i,j}(K^{-1}V_{\psi}K)_{ii}\frac{\rho_j(1-\rho_i)}{1-\rho_i\rho_j}.$$

Next we get

$$tr\{\psi'\Sigma^{-1}\sum_{h=0}^{\infty}\gamma(h+1)'\Sigma^{-1}\psi_h)\}$$
$$= \sum_{h=0}^{\infty} tr\{Q\Omega\alpha_{\perp}(\alpha_{\perp}'\Omega\alpha_{\perp})^{-1}\alpha_{\perp}'\Omega Q'(I_{n_y} - P')^{-1}\Sigma^{-1}P^{2h+1}\}$$
$$= tr\{(I_{n_y} - P)V_{\psi}P(I_{n_y} - P^2)^{-1}\} = tr\{V_{\psi}P(I_{n_y} + P)^{-1}\},$$

with an equivalent expression

$$tr\{V_{\psi}P(I_{n_y} + P)^{-1}\} = \sum_i (K^{-1}V_{\psi}K)_{ii}\frac{\rho_i}{1 + \rho_i}.$$

This shows that $c_3 = tr\{V\}$ is as given the Corollary 2 and we find the additional expression

$$c_3 = \sum_{i,j}(K^{-1}V_{\psi}K)_{ii}\frac{\rho_j(1-\rho_i)}{1-\rho_i\rho_j} + \sum_i (K^{-1}V_{\psi}K)_{ii}\frac{\rho_i}{1+\rho_i}.$$

5   Discussion of the principal steps in the proof

This section contains the basic ideas in the evaluation of the expectation of the likelihood ratio test statistic and hence the ideas of 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) and simplify the derivations. Then we introduce a convenient reparametrization of model $\mathcal{M}_1$ and use it to derive an expansion of the test statistic in Theorem 3. The detailed evaluation of the terms in the expansion are left to the Appendix. We focus on the case $D_t = t^{n_d}$, and mention when necessary the modifications for $D_t = 0$.

We define model $\mathcal{M}_3$ by specifying the cointegration space $sp(\beta^0)$ or equivalently by the restrictions

$$\beta = \beta^0 \tau, \rho = \rho^0 \tau, \ \tau(r \times r).$$

In this case the model equations are

$$\mathcal{M}_3 : \Delta X_t = \alpha \tau (\beta^{0\prime} X_{t-1} + \rho^{0\prime} D_t) + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi d_t + \varepsilon_t. \qquad (20)$$

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 θ denote the parameter and Θ_1, Θ_2, and Θ_3 the parameter sets. We then get

$$\frac{\max_{\theta \in \Theta_2} L(\theta)}{\max_{\theta \in \Theta_1} L(\theta)} = \frac{\max_{\theta \in \Theta_3} L(\theta)}{\max_{\theta \in \Theta_1} L(\theta)} \Big/ \frac{\max_{\theta \in \Theta_3} L(\theta)}{\max_{\theta \in \Theta_2} L(\theta)}$$

or

$$LR(\mathcal{M}_2 | \mathcal{M}_1) = LR(\mathcal{M}_3 | \mathcal{M}_1) / LR(\mathcal{M}_3 | \mathcal{M}_2)$$

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. In the present paper 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 (20) is a simple regression, which facilitates the calculations, and introducing $\mathcal{M}_3$ avoids an expansion of the eigenvalues entering into the trace statistic, since we can use previous results on test for β.

5.1 A reparametrization 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 (20). We find with an obvious notation

$$-2 \log LR(\mathcal{M}_3|\mathcal{M}_1) = -T \log \frac{|M_{\varepsilon\varepsilon.x, \Delta x, D, d}|}{|M_{\varepsilon\varepsilon.\beta^{0\prime}x + \rho^{0\prime}D, \Delta x, d}|}. \tag{21}$$

We want to calculate the expectation of (21) 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 parametrization and regressors using the true value of the parameters, as in Johansen (2000).

We use the notation

$$\Psi^0 = (\Gamma^0_1, \dots, \Gamma^0_{k-1}), \ \Gamma^0 = I_n - \sum_{i=1}^{k-1} \Gamma^0_i,$$

and note that under the null, where $\mathrm{rank}(\alpha^0) = \mathrm{rank}(\beta^0) = r$, the process is $I(1)$ if $|\alpha^{0\prime}_\perp \Gamma^0 \beta^0_\perp| \neq 0$, see Johansen (1996, Theorem 4.2). We let

$$C^0 = \beta^0_\perp (\alpha^{0\prime}_\perp \Gamma^0 \beta^0_\perp)^{-1} \alpha^{0\prime}_\perp,$$

and give Granger's representation

$$X_t = C^0 \sum_{i=1}^{t} (\varepsilon_i + \Phi^0 d_i) + \sum_{i=0}^{\infty} C^0_i (\varepsilon_{t-i} + \Phi^0 d_{t-i} + \alpha^0 \rho^{0\prime} D_{t-i}) + K,$$

where K depends on initial conditions and $\beta^{0\prime} K = 0$. We find from

$$\Pi = \Pi(I - C^0\Gamma^0) + \Pi C^0 \Gamma^0 = \Pi(I_n - C^0\Gamma^0)\bar{\beta}^0 \beta^{0\prime} + \Pi\beta^0_\perp (\alpha^{0\prime}_\perp \Gamma^0 \beta^0_\perp)^{-1} \alpha^{0\prime}_\perp \Gamma^0,$$

since $(I_n - C^0\Gamma^0)\beta_\perp^0 = 0$, that we can choose new parameters as functions of the old

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

The old parameters in terms of the new are given by

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

The null hypothesis $\mathcal{M}_3$ is expressed as $\delta = 0$, which is clearly equivalent to $\Pi = \psi'\beta^{0\prime}$ and $\rho' = \psi'\rho^{0\prime}$.

Model equation (1) with the new parameters is

$$
\Delta X_t = \psi'\beta_e^{0\prime}X_{et-1} + \delta_1'(\alpha_\perp^{0\prime}\Gamma^0\beta_\perp^0)^{-1}\alpha_\perp^{0\prime}\Gamma^0 X_{t-1} + \delta_2' D_t \\
+ \sum_{i=1}^{k-1}\Gamma_i\Delta X_{t-i} + \Phi d_t + \varepsilon_t, \tag{22}
$$

where $\beta_e^{0\prime}X_{et-1} = \beta^{0\prime}X_{t-1} + \rho^{0\prime}D_t$.

Under the null hypothesis $\mathcal{M}_3$, where the rank is r, and β^0 and ρ^0 are known, it holds that $\beta^{0\prime}X_{t-1} + \rho^{0\prime}D_t$ and ΔX_t have a mean that is linear in d_t. Since Φ enters unrestricted, we can replace the regressors $\beta^{0\prime}X_{t-1} + \rho^{0\prime}D_t$ and the lagged differences with the stationary regressors

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

We also want to replace the regressors $(\alpha_\perp^{0\prime}\Gamma^0\beta_\perp^0)^{-1}\alpha_\perp^{0\prime}\Gamma^0 X_{t-1}$ and D_t by something simpler without changing the statistical model and hence the test that $\delta = 0$. We find by summing equation (2) that

$$
\alpha_\perp^{0\prime}(X_t - X_0) = \alpha_\perp^{0\prime}\sum_{i=1}^{k-1}\Gamma_i^0(X_{t-i} - X_{-i}) + \alpha_\perp^{0\prime}\sum_{i=1}^{t}(\varepsilon_i + \Phi^0 d_i).
$$

By subtracting $\sum_{i=1}^{k-1}\Gamma_i^0 X_t$ on both sides and replacing t by $t - 1$ we get

$$
\alpha_\perp^{0\prime}\Gamma^0 X_{t-1} = \alpha_\perp^{0\prime}X_0 + \alpha_\perp^{0\prime}\sum_{i=1}^{k-1}\Gamma_i^0(X_{t-i-1} - X_{t-1} - X_{-i}) + \alpha_\perp^{0\prime}\sum_{i=1}^{t-1}(\varepsilon_i + \Phi^0 d_i).
$$

Because we are correcting for lagged differences in the regression (22) we can replace $(\alpha_\perp^{0\prime}\Gamma^0\beta_\perp^0)^{-1}\alpha_\perp^{0\prime}\Gamma^0 X_{t-1}$ and D_t by the non stationary regressor

$$\left(\begin{array}{c} K_0 + (\alpha_\perp^{0\prime}\Omega\alpha_\perp^0)^{-1/2}\alpha_\perp^{0\prime}\sum_{i=1}^{t-1}(\varepsilon_i + \Phi^0 d_i) \\ D_t \end{array} \middle| d_t \right),$$

of dimension $n_a = n - r + 1$, and where K_0 depends on initial conditions. A non-singular linear transformation of this yields the regressor

$$A_{t-1} = \left(\begin{array}{c} (\alpha_\perp^{0\prime}\Omega^0\alpha_\perp^0)^{-1/2}\alpha_\perp^{0\prime}\sum_{i=1}^{t-1}\varepsilon_i \\ D_t \end{array} \middle| d_t \right) = \left(\begin{array}{c} \sum_{i=1}^{t-1} B_i \\ D_t \end{array} \middle| d_t \right). \qquad (23)$$

Obviously if $D_t = 0$, we do not extend the process. We define $\mu = n_d^{-1}(\alpha_\perp^{0\prime}\Omega^0\alpha_\perp^0)^{-1/2}\alpha_\perp^{0\prime}\Phi_{n_d-1}^0$ and $A_{t-1} = \left(\sum_{i=1}^{t-1} B_i + \mu t^{n_d} \middle| d_t\right)$ of dimension $n_a = n - r$. Model equation (22) in the new variables and with suitably redefined parameters Ψ and Φ becomes

$$\underset{(n)}{\Delta X_t} = \underset{(r)}{\psi' V_{t-1}} + \underset{(n_a)}{\delta' A_{t-1}} + \underset{((k-1)n)}{\Psi\ Z_{t-1}} + \underset{(n_d)}{\Phi\ d_t} + \underset{(n)}{\varepsilon_t}, \qquad (24)$$

where the dimensions are indicated below each variable. The test for $\mathcal{M}_3$ in $\mathcal{M}_1$ is the test for $\delta = 0$ in (24). 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 $\mathcal{M}_2$ can be reparametrized as

$$\underset{(n)}{\Delta X_t} = \underset{(r)}{\alpha V_{t-1}} + \underset{(n_a)}{\alpha\delta' A_{t-1}} + \underset{((k-1)n)}{\Psi\ Z_{t-1}} + \underset{(n_d)}{\Phi\ d_t} + \underset{(n)}{\varepsilon_t}, \qquad (25)$$

see equation (14) in Johansen (2000), and the test for $\mathcal{M}_3$ in $\mathcal{M}_1$ is the test for $\delta = 0$ in (25).

This formulation covers both the case $D_t = t^{n_d}$ and $D_t = 0$.

5.2   The likelihood ratio test and its expansion

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

$$\sum_{t=1}^{T} \left(\begin{array}{c} \Delta X_t \\ V_{t-1} \\ \varepsilon_t \end{array} \right) \left(\begin{array}{c} \Delta X_t \\ V_{t-1} \\ \varepsilon_t \end{array} \right)' = \left(\begin{array}{ccc} M_{00} & M_{0v} & M_{0\varepsilon} \\ M_{v0} & M_{vv} & M_{v\varepsilon} \\ M_{\varepsilon 0} & M_{\varepsilon v} & M_{\varepsilon\varepsilon} \end{array} \right).$$

We use the notation for any three process X_t, U_t, and V_{t-1}, say,

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

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)

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

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

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

which is just another expression for (21), but given in terms of processes that are normalized. Hence with $Q = TS_{\varepsilon\varepsilon.v}^{-1}S_{\varepsilon a.v}S_{aa.v}^{-1}S_{a\varepsilon.v}$ we find

$$-2\log LR(\mathcal{M}_3|\mathcal{M}_1) = -T\log\left|I_n - T^{-1}Q\right|,$$

which implies

$$-2\log LR(\mathcal{M}_3|\mathcal{M}_1) \overset{1}{=} tr\{Q\} + \frac{1}{2T}tr\{Q^2\}. \tag{26}$$

We use this expression derived from (24) combined with a similar expression from Johansen (2000) for $-2\log LR(\mathcal{M}_3|\mathcal{M}_2)$ derived from (25) to find an expansion for $-2\log LR(\mathcal{M}_2|\mathcal{M}_1)$ which will form the basis for the further calculations. We define

$$\kappa = (\alpha'\Omega^{-1}\alpha)^{-\frac{1}{2}}, \Sigma_{\beta\beta} = \mathrm{Var}(\beta'X_t|\Delta X_t',\ldots,\Delta X_{t-k+1}').$$

Theorem 3 *The likelihood ratio test for cointegrating rank in model (1) has the expansion*

$$\begin{aligned}
&-2\log LR(\mathcal{M}_2|\mathcal{M}_1)\\
\overset{1}{=}\quad &T tr\{S_{ab.v}S_{bb.v}^{-1}S_{ba.v}S_{aa.v}^{-1}\} + \tfrac{1}{2}T tr\{(S_{ab.v}S_{bb.v}^{-1}S_{ba.v}S_{aa.v}^{-1})^2\}\\
-\quad &2tr\{S_{ba.v}S_{aa}^{-1}S_{au.v,b}\kappa'S_{vv}^{-1}S_{vb}\} - T^{-1}tr\{\kappa\Sigma_{\beta\beta}^{-1}\kappa S_{ua}S_{aa}^{-1}S_{ab}S_{ba}S_{aa}^{-1}S_{au}\}.\\
-\quad &tr\{S_{ba}S_{aa}^{-1}S_{ab}S_{bv}S_{vv}^{-1}\kappa\kappa'S_{vv}^{-1}S_{vb}\} + tr\{S_{ua}S_{aa}^{-1}S_{au}\kappa'S_{vv}^{-1}S_{vb}S_{bv}S_{vv}^{-1}\kappa\}\\
+\quad &2tr\{S_{ba}S_{aa}^{-1}S_{au}\kappa'S_{vv}^{-1}S_{vu}\kappa'S_{vv}^{-1}S_{vb}\}.
\end{aligned}$$

The proof will be given in the Appendix. By a detailed analysis of these terms one can then find an expression for the expectation of $-2\log LR(\mathcal{M}_2|\mathcal{M}_1)$ and hence prove Theorem 1 and hence Corollary 2.

6   Conclusion

A detailed analysis of the Taylor's expansion of the trace statistic gives an approximation of its expectation, which is used for a correction factor of the form

$$a(T, n_b, n_d)(1 + T^{-1}b(n_b, n_d, \theta)).$$

A numerical approximation to $a(T, n_b, n_d)$ is found by simulation and a computable formula for $b(n_b, n_d, \theta)$ is given.

A general conclusion from the simulation experiments is that as θ approaches a boundary point, where the cointegration and integration properties change, the size of a nominal 5% test tends to one, and the corrected size tends to zero, since the correction factor has a singularity at that point.

Throughout, however, the corrected size is closer to the nominal value, so there seems to be an 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.

7   Acknowledgment

8   References

Abadir, K. M., Hadri K., and Tzavalis, E. (1999), The influence of VAR dimension on estimator biases, *Econometrica* 67, 163-181.

Ahn, S. K. and Reinsel, C. G. (1990), Estimation for partially non-stationary 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, *Proceeding of the Royal Society of London* Series A, Vol. 160, 268-282.

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

Bravo, F. (1998), A correction factor for unit root test statistics, Discussion paper No. 9809, University of Southampton.

Cheung, Y-W, and Lai, K. S. (1993), Finite sample sizes of Johansens likelihood ratio tests for cointegration, *Oxford Bulletin of Economics and Statistics* 55, 313- 32.

Gonzalo, J. and Pitarakis, J.-Y. (1999), Dimensionality effect in cointegration analysis, in *Cointegration, Causality and Forecasting: Festschrift in Honour of Clive Granger*, R. F. Engle and H. White (eds.), Oxford University Press, Oxford.

Hansen, H. and Rahbek, A. C. (2000), Approximate conditional unit root inference, Discussion paper, University of Copenhagen.

Haug, A. A. (1996), Tests for cointegration: A Monte Carlo comparison, *Journal of Econometrics* 71, 89-115

Jensen, J. L and Wood, A. T. A. (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.

Johansen, S. (1996), *Likelihood-based inference in cointegrated vector autoregressive models,* Oxford University Press, Oxford.

Johansen, S. (2000), The Bartlett correction of the test for rank in the cointegrated vector autoregressive model, (forthcoming *Econometric Theory)*

Larsson, R. (1998a), Bartlett corrections for unit root test statistics, *Journal of Times Series Analysis* 19, 426-438.

Larsson, R. (1998b), Distribution approximation of unit root test in autoregressive models, *Econometrics Journal* 1, 10-26.

Lawley, D. N. (1956), A general method for approximating to 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.

Nielsen, B. (1997b), On the distribution of tests for cointegration rank, Discussion Paper No. 113, Nuffield College, Oxford.

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.

9   Appendix

9.1   Proof of Theorem 3

Proof. We start by expanding the matrix Q, see (26), by introducing the variables U_t and B_t, see (5) and (6), and find

$$
\begin{aligned}
tr\{Q\} = \ & Ttr\{S_{a\varepsilon.v}S_{\varepsilon\varepsilon.v}^{-1}S_{\varepsilon a.v}S_{aa.v}^{-1}\} \\
= \ & Ttr\{\begin{pmatrix} S_{ua.v} \\ S_{ba.v} \end{pmatrix}' \begin{pmatrix} S_{uu.v} & S_{ub.v} \\ S_{bu.v} & S_{bb.v} \end{pmatrix}^{-1} \begin{pmatrix} S_{ua.v} \\ S_{ba.v} \end{pmatrix} S_{aa.v}^{-1}\} \\
= \ & Ttr\{S_{ab.v}S_{bb.v}^{-1}S_{ba.v}S_{aa.v}^{-1}\} + Ttr\{S_{au.v,b}S_{uu.v,b}^{-1}S_{ua.v,b}S_{aa.v}^{-1}\} \\
= \ & tr\{Q_1 + Q_2\},
\end{aligned}
$$

such that from (26) we find

$$
-2\log LR(\mathcal{M}_3|\mathcal{M}_1) \stackrel{1}{=} tr\{Q_1 + Q_2\} + \frac{1}{2}T^{-1}tr\{(Q_1 + Q_2)^2\}. \qquad (27)
$$

The term Q_2 can be rewritten as follows

$$
\begin{aligned}
Q_2 = \ & TS_{au.v,b}S_{uu.v,b}^{-1}S_{ua.v,b}S_{aa.v}^{-1} \\
= \ & TS_{au.v,b}S_{uu.v,b}^{-1}S_{ua.v,b}S_{aa.v,b}^{-1} + TS_{au.v,b}S_{uu.v,b}^{-1}S_{ua.v,b}(S_{aa.v}^{-1} - S_{aa.v,b}^{-1}), \qquad (28) \\
= \ & Q_{21} + Q_{22},
\end{aligned}
$$

where $S_{aa.v,b} = S_{aa.v} - S_{ab.v}S_{bb.v}^{-1}S_{ba.v}$, such that

$$
S_{aa.v,b}^{-1} = S_{aa.v}^{-1} + S_{aa.v}^{-1}S_{ab.v}S_{bb.v}^{-1}S_{ba.v}S_{aa.v}^{-1} + O_P(S_{aa}^{-1}T^{-2}).
$$

Hence the term, Q_{22}, in (28) becomes

$$
\begin{aligned}
Q_{22} =&\ T S_{au.v,b} S_{uu.v,b}^{-1} S_{ua.v,b} (S_{aa.v}^{-1} - S_{aa.v,b}^{-1}) \\
\overset{1}{=}&\ -T S_{au.v,b} S_{uu.v,b}^{-1} S_{ua.v,b} S_{aa.v}^{-1} S_{ab.v} S_{bb.v}^{-1} S_{ba.v} S_{aa.v}^{-1} \\
=&\ -T^{-1} Q_2 Q_1,
\end{aligned}
$$

and for Q_2^2 we find

$$
Q_2^2 \overset{0}{=} (S_{au} S_{ua} S_{aa}^{-1})^2.
$$

Thus

$$
-2\log LR(\mathcal{M}_3|\mathcal{M}_1) \overset{1}{=} tr\{Q_1\} + tr\{Q_{21}\} + \tfrac{1}{2} T^{-1} tr\{Q_1^2\} + \tfrac{1}{2} T^{-1} tr\{Q_2^2\}
$$
$$
\overset{1}{=} T tr\{S_{ab.v} S_{bb.v}^{-1} S_{ba.v} S_{aa.v}^{-1}\} + T tr\{S_{aa.v,b}^{-1} S_{au.v,b} S_{uu.v,b}^{-1} S_{ua.v,b}\}
$$
$$
+\tfrac{1}{2} T tr\{(S_{ab.v} S_{bb.v}^{-1} S_{ba.v} S_{aa.v}^{-1})^2\} + \tfrac{1}{2} T^{-1} tr\{(S_{au} S_{ua} S_{aa}^{-1})^2\}.
$$

(29)

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

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

Subtracting this result from (29), we have finished the proof of the expansion in Theorem 3. ∎

9.2 Asymptotic moments

The next Lemma contains the results about product moments that can be used to evaluate the various terms in the expansion of the likelihood ratio test. We define the standard Brownian motion $W(s)$ by the limit

$$
T^{-\frac{1}{2}} \sum_{i=1}^{[Ts]} B_i \overset{w}{\to} W(s)
$$

of dimension $n_b = n - r$. Suitably normalized the process $A_{[Ts]}$, see (23), converges to a limit $F(s)$, which depends on $W(s)$ and the deterministic terms. The limit of the moments can be expressed in terms of W and F.

Lemma 4 *Let $D_t = t^{n_d}$, $\Phi d_t = \sum_{i=0}^{n_d-1} \Phi_i t^i$ $n_d = 0, 1, 2$. Let A_{t-1} be given by (23), then*

$$\begin{pmatrix} T^{-1/2} I_{n_b} & 0 \\ 0 & T^{-n_d} \end{pmatrix} A_{[Ts]} \xrightarrow{w} F(s) = \begin{pmatrix} W(s) \\ s^{n_d} \end{pmatrix} \Big| 1, \ldots, s^{n_d-1} \Big). \qquad (30)$$

It follows that

$$\begin{aligned}
M_1 &\xrightarrow{w} \int_0^1 (dW) F' \left(\int_0^1 FF' ds \right)^{-1} \int_0^1 F(dW)' = M_1(n_b, n_d), \\
M_3 &\xrightarrow{w} \int_0^1 (dW) F' \left(\int_0^1 FF' ds \right)^{-1} \left(\int_0^1 F(dW)' + \Upsilon \right) = M_3(n_b, n_d), \\
M_6 &\xrightarrow{w} \left(\Upsilon' + \int_0^1 (dW) F' \right) \left(\int_0^1 FF' ds \right)^{-1} \left(\int_0^1 F(dW)' + \Upsilon \right) = M_6(n_b, n_d), \\
\Upsilon' &= (I_{n_b}, 0).
\end{aligned}$$

$$\qquad (31)$$

The matrices

$$\begin{aligned}
M_2(n_b, n_d) &= M_1^2(n_b, n_d) \\
M_4(n_b, n_d) &= M_3'(n_b, n_d) M_3(n_b, n_d)
\end{aligned}$$

can be expressed similarly. For $i = 1, \ldots, 6$, the matrices $\lim_T E[M_i]$ are proportional to the identity matrix such that for any matrix K we have

$$E[tr\{M_i K\}] \to m_i(n_b, n_d) n_b^{-1} tr\{K\}, \qquad (32)$$

where we define the coefficients $m_i(n_b, n_d) = \lim_T E[tr\{M_i\}]$.

Finally the relation

$$E[M_4 - 2M_5 - n_b M_6] \to 0, \qquad (33)$$

has been verified by simulation for $n_d = 0, 1, 2$.

If $d_t = D_t = 0$, these results hold with $F = W$, and $\Upsilon = I_{n_b}$.

Proof. The results (30) and (31) follow from standard results about Brownian motion. Note that M_5 is different from the other moments. It has a constant mean but a variance that grows like T, which makes (33) difficult and time consuming to check.

To prove (32) we let $H(W)$ be any of the limit functionals given in (31). It is not difficult to see that for any orthonormal matrix O :

$$O' H(W) O = H(O'W),$$

which has the same distribution as $H(W)$, and hence the same expectation. Thus $E[H(W)]$ is invariant under orthogonal transformations and hence proportional to the identity matrix I_{n_b}. The elements in the diagonal can be expressed as $tr\{E[H(W)]\}n_b^{-1}$. $\blacksquare$

We next define the processes

$$
\begin{aligned}
T^{-\frac{1}{2}} \sum_{t=1}^{[Ts]} U_t &\xrightarrow{w} W^u(s) \\
T^{-\frac{1}{2}} \sum_{t=1}^{[Ts]} (B_t B_t' - I_{n_b}) &\xrightarrow{w} W^{bb}(s), \\
T^{-\frac{1}{2}} \sum_{t=1}^{[Ts]} B_t B_{t-i}' &\xrightarrow{w} W_i^{bb}(s), i = 1, \dots, \\
T^{-\frac{1}{2}} \sum_{t=1}^{[Ts]} B_t U_{t-i}' &\xrightarrow{w} W_i^{bu}(s), i = 1, \dots
\end{aligned}
\tag{34}
$$

It is seen that $W^b = W, W^u, W^{bb}, W_1^{bu}, W_1^{bb}, \dots$ are all mutually independent and hence that $W^{bb}, W_1^{bu}, W_1^{bb}, \dots$, are independent of the limits involving W^b and W^u.

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

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

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

Lemma 5

$$
T^{-1} E(M_{by} M_{yb}) \to tr\{\Sigma\} I_{n_b}
\tag{35}
$$

$$
T^{-1} E(M_{yb} M_{by}) \to n_b \Sigma
\tag{36}
$$

$$
E[\theta M_{ua}^+ + \psi M_{ba}^+) M_{aa}^{-1} (M_{au}^+ \theta' + M_{ab}^+ \psi')] \overset{0}{=} n_a \theta \theta' + \psi E[M_{ba}^+ M_{aa}^{-1} M_{ab}^+] \psi'
\tag{37}
$$

$$
E[M_{aa}^{-1} (M_{au}^+ \theta' + M_{ab}^+ \psi')(\theta M_{ua}^+ + \psi M_{ba}^+)] \overset{0}{=} tr\{\theta' \theta\} I_{n_a} + E[M_{aa}^{-1} M_{ab}^+ \psi' \psi M_{ba}^+]
\tag{38}
$$

$$
E[T^{-1} (M_{yy} - T I_{n_y}) M_{yb}] \to \sum_{k=0}^{\infty} \gamma(k+1)' \psi_k + \psi_k tr\{\gamma(k+1)'\}
\tag{39}
$$

$$
T^{-1/2}((M_{yy} - T I_{n_y}), M_{yb}) \text{ asymp. indep. of } M_{ba} M_{aa}^{-1} M_{ay}
\tag{40}
$$

$$
M_{aa}^{-1/2} M_{ay} \text{ asymp. distr. as } M_{aa}^{-1/2}(M_{au}^+ \theta' + M_{ab}^+ \psi')
\tag{41}
$$

Proof. Proof of (35):

$$T^{-1}E(M_{by}M_{yb})$$
$$= T^{-1}\sum_{s,t,i,j} E[B_t(U'_{t-i-1}\theta'_i + B'_{t-i-1}\psi'_i)(\theta_j U_{s-j-1} + \psi_j B_{s-j-1})B'_s]$$
$$= T^{-1}\sum_{t,i} E[B_t U'_{t-i-1}\theta'_i\theta_i U_{t-i-1}B'_t] + T^{-1}\sum_{t,i} E[B_t B'_{t-i-1}\psi'_i\psi_i B_{t-i-1}B'_t]$$
$$\overset{0}{=} (\sum_{i=0}^{\infty} tr\{\theta'_i\theta_i\} + tr\{\psi'_i\psi_i\})I_{n_b} = tr\{\Sigma\}I_{n_b}.$$

Proof of (36):

$$T^{-1}E(M_{yb}M_{by})$$
$$= T^{-1}\sum_{s,t,i,j} E[(\theta_j U_{s-j-1} + \psi_j B_{s-j-1})B'_s B_t(U'_{t-i-1}\theta'_i + B'_{t-i-1}\psi'_i)]$$
$$= T^{-1}\sum_{t,i} E[\theta_i U_{t-i-1}B'_t B_t U'_{t-i-1}\theta'_i] + T^{-1}\sum_{t,i} E[\psi_i B_{t-i-1}B'_t B_t B'_{t-i-1}\psi'_i]$$
$$\overset{0}{=} n_b \sum_{i=0}^{\infty} (\theta_i\theta'_i + \psi_i\psi'_i) = n_b\Sigma.$$

The results (37) and (38) follow by similar calculations.

Proof of (39): We find

$$T^{-1}E[(M_{yy} - TI_{n_y})M_{yb}]$$
$$= T^{-1}E[\sum_{s,l,j,k,m}\{\theta_j(U_{s-j-1}U'_{s-k-1} - \delta_{jk}I_r)\theta'_k + \psi_j B_{s-j-1}U'_{s-k-1}\theta'_k$$
$$+ \theta_j U_{s-j-1}B'_{s-k-1}\psi'_k + \psi_j(B_{s-j-1}B'_{s-k-1} - \delta_{jk}I_r)\psi'_k\}$$
$$\times(\theta_m U_{l-m-1} + \psi_m B_{l-m-1})B'_l]\}.$$

We get three contributions A_1, A_2, and A_3. First

$$A_1 = T^{-1}E[\sum_{s,l,j,k,m}\psi_j B_{s-j-1}U'_{s-k-1}\theta'_k\theta_m U_{l-m-1}B'_l].$$

This gives a contribution if $s - j - 1 = l$ and $s - k - 1 = l - m - 1$, which implies $k = j + m + 1$, and then we find

$$A_1 \overset{0}{=} \sum_{j,m} tr\{\theta'_{j+m+1}\theta_m\}\psi_j = \sum_{j=0}^{\infty}\psi_j tr\{\gamma_\theta(j+1)'\}.$$

Next

$$A_2 = T^{-1}E[\sum_{s,l,k,j,m}\theta_j U_{s-j-1}B'_{s-k-1}\psi'_k\theta_m U_{l-m-1}B'_l]$$
$$= T^{-1}E[\sum_{s,l,k,j,m}\theta_j U_{s-j-1}U'_{l-m-1}\theta'_m\psi_k B_{s-k-1}B'_l].$$

Here we get a contribution if $s - j - 1 = l - m - 1$, and $l = s - k - 1$, which implies that $j = k + 1 + m$, and hence

$$A_2 \overset{0}{=} \sum_{j,m}\theta_{j+m+1}\theta'_m\psi_j = \sum_{j=0}^{\infty}\gamma_\theta(j+1)'\psi_j.$$

27

Finally we get the term

$$A_3 \stackrel{0}{=} T^{-1}E[\sum_{s,l,j,k,m} \psi_j(B_{s-j-1}B'_{s-k-1} - \delta_{jk}I_r)\psi'_k\psi_m B_{l-m-1}B'_l].$$

Here we get a contribution for $s - j - 1 = l - m - 1$ and $s - k - 1 = l$, which implies $j = k + 1 + m$ and for $s - j - 1 = l$ and $s - k - 1 = l - m - 1$, which implies $k = j + m + 1$. Hence

$$
\begin{aligned}
A_3 = & \ T^{-1}E[\sum_{s,l,k,m} \psi_{k+m+1}B_{l-m-1}B'_l\psi'_k\psi_m B_{l-m-1}B'_l] \\
& +T^{-1}E[\sum_{s,l,j,m} \psi_j B_l B'_{l-m-1}\psi'_{j+m+1}\psi_m B_{l-m-1}B'_l] \\
\stackrel{0}{=} & \ \sum_{k,m} \psi_{k+m+1}\psi'_m\psi_k + \sum_{j,m} \psi_j tr\{\psi'_{j+m+1}\psi_m\} \\
= & \ \sum_{k=0}^{\infty} \gamma_\psi(k+1)'\psi_k + \sum_{j=0}^{\infty} \psi_j tr\{\gamma_\psi(j+1)'\}.
\end{aligned}
$$

Adding the contributions from A_1, A_2, and A_3, we find the result (39).

Proof of (40): The limit of $T^{-1/2}(M_{yy} - TI_{n_y})$ and $T^{-1/2}M_{yb}$ involve the Brownian motions W^{bb} and W^{bu}, whereas the limit of $M_{ba}M_{aa}^{-1}M_{ay}$ involves W^b and W^u, see (34).

Proof of (41): The limit of $M_{aa}^{-1/2}M_{ay}$ is given by

$$
\begin{aligned}
M_{aa}^{-1/2}M_{ay} &= M_{aa}^{-1/2}\sum_{i=0}^{\infty}\sum_t A_{t-1}(U'_{t-i-1}\theta'_i + B'_{t-i-1}\psi_i) \\
&\stackrel{w}{\rightarrow} [\int_0^1 FF'du]^{-1/2}[\int_0^1 Fd(W^u)'\theta' + [\int_0^1 Fd(W^b)' + \Upsilon]\psi'],
\end{aligned}
$$

which is the same limit as $M_{aa}^{-1/2}(M_{au}^+\theta' + M_{ab}^+\psi')$. ∎

Lemma 6

$$M_{bd} \ \text{asymp. indep. of} \ (M_{ba}M_{aa}^{-1}M_{ab}^+, M_{ba}M_{aa}^{-1}M_{ab})$$

Proof. Consider the statistical model for the parameters $v_0, \ldots, v_{n_d-1}$:

$$dZ_s = \sum_{i=0}^{n_d-1} v_i s^i ds + dW_s, 0 \le s \le 1, \tag{42}$$

where W_s is standard Brownian motion. We define the statistics $S_i = \int_0^1 s^i dW(s)$, $i = 0, \ldots, n_d - 1$, and

$$
\begin{aligned}
H &= \int_0^1 (dZ)F'\left(\int_0^1 FF'ds\right)^{-1}\int_0^1 F(dZ)', \\
H^+ &= \int_0^1 (dZ)F'\left(\int_0^1 FF'ds\right)^{-1}(\int_0^1 F(dZ)' + \Upsilon).
\end{aligned}
$$

with

$$F = \left(\begin{array}{c} Z_s \\ s^{n_d} \end{array} \,\middle|\, 1, \ldots, s^{n_d-1} \right).$$

The statistic $S = (S_0, \ldots, S_{n_d-1})$ is multivariate Gaussian and minimal sufficient and complete for the parameters $v_0, \ldots, v_{n_d-1}$. We next want to prove that H is ancillary, and it follows that it is independent of S by Basu's theorem, see Basu (1955).

By integrating (42) it is seen that the

$$Z_s = Z_0 + \sum_{i=0}^{n_d-1} \frac{v_i}{i+1} s^{i+1} + W_s, 0 \le s \le 1,$$

hence

$$F = \left(\begin{array}{c} Z_0 + v_{n_d-1} n_d^{-1} s^{n_d} + W_s \\ s^{n_d} \end{array} \,\middle|\, 1, \ldots, s^{n_d-1} \right)$$

$$= \left(\begin{array}{cc} 1 & v_{n_d-1} n_d^{-1} + Z_0 1_{\{n_d=0\}} \\ 0 & 1 \end{array} \right) \left(\begin{array}{c} W_s \\ s^{n_d} \end{array} \,\middle|\, 1, \ldots, s^{n_d-1} \right).$$

The statistic H is invariant under the linear transformation, such that all parameters can be eliminated from the expressions involving F, and F can be replaced by

$$F = \left(\begin{array}{c} W_s \\ s^{n_d} \end{array} \,\middle|\, 1, \ldots, s^{n_d-1} \right).$$

Finally it is seen that from (42) that

$$\int_0^\infty (dZ_s) F' = \sum_{i=0}^{n_d-1} v_i \int_0^\infty s^i F' ds + \int_0^1 (dW_s) F' = \int_0^1 (dW_s) F',$$

since F is orthogonal to $1, \ldots, s^{n_d-1}$. Thus the statistic H is ancillary and hence independent of $S_0, , \ldots, S_{n_d-1}$. The same proof can be used for H^+.

The limit of M_{bd} is a function of $S_0, , \ldots, S_{n_d-1}$, whereas the limit of $(M_{ba} M_{aa}^{-1} M_{ab}^+, M_{ba} M_{aa}^{-1} M_{ab})$ can be expressed in terms of H and H^+. ∎

9.3   Proof of Theorem 1

Proof. We apply the expansion from Theorem 3, and we find that the last five terms are calculated in Johansen (2000, proof of Theorem 4). It is shown there

that the total contribution of these terms is

$$-(n-r)(n-r+1)v_\theta, \tag{43}$$

where $v_\theta = tr\{\theta'\Sigma^{-1}\theta\}$, see (8). What remains are the two terms

$$TE[tr\{S_{ab.v}S_{bb.v}^{-1}S_{ba.v}S_{aa.v}^{-1}\}] + \tfrac{1}{2}TE[tr\{(S_{ab.v}S_{bb.v}^{-1}S_{ba.v}S_{aa.v}^{-1})^2\}]. \tag{44}$$

The second term in (44) is easy because of the factor T^{-1}, and we find the evaluations

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

such that

$$\tfrac{1}{2}TE[tr\{(S_{ab.v}S_{bb.v}^{-1}S_{ba.v}S_{aa.v}^{-1})^2\}] \overset{1}{=} \tfrac{1}{2}TE[tr\{(M_{bb.d}^{-1}M_{ba}M_{aa}^{-1}M_{ab})^2\}]. \tag{45}$$

What remains is the first term of (44). We expand it as follows:

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

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}:

$$\begin{aligned}
&E[Ttr\{S_{ab.v}S_{bb.v}^{-1}S_{ba.v}S_{aa.v}^{-1}\}]\\
&\overset{1}{=} TM_{ab}M_{bb.d}^{-1}M_{ba}M_{aa}^{-1}\\
&-T(M_{ay}M_{yy.d}^{-1}M_{yb.d})M_{bb.d}^{-1}M_{ba}M_{aa}^{-1} - TM_{ab}M_{bb.d}^{-1}(M_{by.d}M_{yy.d}^{-1}M_{ya})M_{aa}^{-1}\\
&+TM_{ab}M_{bb.d}^{-1}(M_{by.d}M_{yy.d}^{-1}M_{yb.d})M_{bb.d}^{-1}M_{ba}M_{aa}^{-1}\\
&+TM_{ab}M_{bb.d}^{-1}M_{ba}M_{aa}^{-1}(M_{ay}M_{yy.d}^{-1}M_{ya})M_{aa}^{-1}\\
&+T(M_{ay}M_{yy.d}^{-1}M_{yb.d})M_{bb.d}^{-1}(M_{by.d}M_{yy.d}^{-1}M_{ya})M_{aa}^{-1}\\
&= K_0 + T^{-1}(K_1 + K_2 + K_3 + K_4).
\end{aligned} \tag{46}$$

The first term is

$$E[K_0] = TE[tr\{M_{ab}M_{bb.d}^{-1}M_{ba}M_{aa}^{-1}\}],$$

which we combine with (45) to the term

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

which is the main term in Theorem 1.

For the remaining terms in (46) we can prove with

$$v_\theta = tr\{V_\theta\}, v_\psi = tr\{V_\psi\}, \tau_\theta = tr\{\gamma_\theta(0)\Sigma^{-1}\}, \tau_\psi = tr\{\gamma_\psi(0)\Sigma^{-1}\}$$

$$
\begin{aligned}
E[K_1] =& \ -2T^2 E[tr\{(M_{ay}M_{yy.d}^{-1}M_{yb.d})M_{bb.d}^{-1}M_{ba}M_{aa}^{-1}\}] \\
\overset{0}{=}& \ tr\{E[M_1]\}(-v_\theta - \tau_\theta + v_\psi - \tau_\psi) + tr\{E[M_3]\}(-v_\theta + \tau_\theta - v_\psi + \tau_\psi) \\
& - 2tr\{E[M_5]\psi'\Sigma^{-1}\psi + n_d E[M_3]\psi'\Sigma^{-1}\psi - E[M_3]V\}, \\
E[K_2] =& \ T^2 E[tr\{M_{ab}M_{bb.d}^{-1}(M_{by.d}M_{yy.d}^{-1}M_{yb.d})M_{bb.d}^{-1}M_{ba}M\}] \\
\overset{0}{=}& \ (r + (k-1)n)tr\{E[M_1]\}, \\
E[K_3] =& \ T^2 E[tr\{M_{ab}M_{bb.d}^{-1}M_{ba}M_{aa}^{-1}(M_{ay}M_{yy.d}^{-1}M_{ya})M_{aa}^{-1}\}] \\
\overset{0}{=}& \ tr\{E[M_1]\}v_\theta + tr\{E[M_4]\psi'\Sigma^{-1}\psi\}, \\
E[K_4] =& \ T^2 E[tr\{M_{ay}M_{yy.d}^{-1}M_{yb.d}M_{bb.d}^{-1}M_{by.d}M_{yy.d}^{-1}M_{ya}M_{aa}^{-1}\}] \\
\overset{0}{=}& \ (n-r+1)(n-r)v_\theta + (n-r)tr\{E[M_6]\psi'\Sigma^{-1}\psi\},
\end{aligned}
$$

where $\psi'\Sigma^{-1}\psi$ is given by (8) and V by (19). Adding these contributions we note that the first term in $E[K_4]$ cancels (43) and find the result in Theorem 1.

The final part of this proof contains a detailed evaluation of the terms $E[K_1], \ldots, E[K_4]$, in order to prove these relations.

K1

Note that we have

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

such that

$$
\begin{aligned}
2tr\{\Sigma^{-1}\textstyle\sum_{h=1}^{\infty} \gamma_\theta(h)\} &= (tr\{V_\theta\} - tr\{\gamma_\theta(0)\}) &&= (v_\theta - \tau_\theta), \\
2tr\{\Sigma^{-1}\textstyle\sum_{h=0}^{\infty} \gamma_\theta(h)\} &= (tr\{V_\theta\} + tr\{\gamma_\theta(0)\}) &&= (v_\theta + \tau_\theta), \\
2tr\{\Sigma^{-1}\textstyle\sum_{h=1}^{\infty} \gamma_\psi(h)\} &= (tr\{V_\psi\} - tr\{\gamma_\psi(0)\}) &&= (v_\psi - \tau_\psi), \\
2tr\{\Sigma^{-1}\textstyle\sum_{h=0}^{\infty} \gamma_\psi(h)\} &= (tr\{V_\psi\} + tr\{\gamma_\psi(0)\}) &&= (v_\psi + \tau_\psi).
\end{aligned}
\tag{47}
$$

We find

$$
\begin{aligned}
K_1 & \\
=& \ -2T^2 tr\{M_{ay}M_{yy.d}^{-1}M_{yb.d}M_{bb.d}^{-1}M_{ba}\} \\
=& \ -2tr\{M_{ay}(I_{n_y} - (I_{n_y} - T^{-1}M_{yy.d}))^{-1}(M_{yb} - M_{yd}M_{dd}^{-1}M_{db}) \\
& \times (I_{n_b} - (I_{n_b} - T^{-1}M_{bb.d}))^{-1}M_{ba}\} \\
=& \ -2tr\{M_{ay}M_{yb}M_{ba}\} + 2T^{-1}tr\{M_{ay}(M_{yy.d} - TI_{n_y})M_{yb}M_{ba}\} \\
& -2tr\{M_{ay}M_{yd}M_{dd}^{-1}M_{db}M_{ba}\} + 2T^{-1}tr\{M_{ay}M_{yb}(M_{bb.d} - TI_{n_b})M_{ba}\} \\
=& \ K_{11} + K_{12} + K_{13} + K_{14}
\end{aligned}
$$

31

In the following we always find the expectation by first conditioning on the permanent shocks B_t. This makes the evaluation of the expectations easier and separates the factors due to the common trends and those due to the transitory shocks. We also normalize on $M_{aa} = I_{n_a}$, and $\Sigma = I_{n_y}$, for notational convenience.

K$_{11}$

$$
\begin{aligned}
& E[K_{11}] \\
= & -2E[tr\{M_{ay}M_{yb}M_{ba}\}] \\
= & -2E[tr\{\textstyle\sum_{t,s,i,j} A_{t-1}(U'_{t-i-1}\theta'_i + B'_{t-i-1}\psi'_i)(\theta_j U_{s-j-1} + \psi_j B_{s-j-1})B'_s M_{ba}\}] \\
= & -2E[tr\{\textstyle\sum_{t,s,i,j} A_{t-1}U'_{t-i-1}\theta'_i\theta_j U_{s-j-1}B'_s M_{ba}\}] \\
& -2E[tr\{\textstyle\sum_{t,s,i,j} A_{t-1}B'_{t-i-1}\psi'_i\psi_j B_{s-j-1}B'_s M_{ba}\}] = K_{111} + K_{112}.
\end{aligned}
$$

In the first term we get a contribution for $t - i - 1 = s - j - 1$, and find

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

where we have replaced $\sum_t A_{t-1}B'_{t-i+j}$ with M_{ab} if $i \leq j$ and M^+_{ab} is $i > j$. We then get, see (47)

$$
\begin{aligned}
E[K_{111}] & \overset{0}{=} -2tr\{E[M_1]\}\textstyle\sum_{h=0}^{\infty} tr\{\gamma_\theta(h)\} - 2tr\{E[M_3]\}\textstyle\sum_{h=1}^{\infty} tr\{\gamma_\theta(h)\} \\
& = -tr\{E[M_1]\}(v_\theta + \tau_\theta) - tr\{E[M_3]\}(v_\theta - \tau_\theta).
\end{aligned}
$$

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

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

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

$$
K_{1121} \overset{0}{=} -2E[tr\{\psi'\sum_j \psi_j (\sum_s B_{s-j-1}B'_s M_{ba}M^+_{ab})\}].
$$

32

Simulations show that

$$\lim_{T \to \infty} E[\sum_s B_{s-j-1} B'_s M_{ba} M^+_{ab}]$$

is independent of $j = 0, 1, \ldots$, and we find,

$$K_{1121} \overset{0}{=} -2tr\{E[M_5]\psi'\psi\}.$$

For the next term we find a contribution for $t - m = s, t - i - 1 = s - j - 1$, such that $i = m + j$, that is $j < i$. We then get

$$
\begin{aligned}
E[K_{1122}] =\ & -2E[tr\{\sum_{t,s,i,j}(\sum_{m=1}^{i}\Delta A_{t-m})B'_{t-i-1}\psi'_i\psi_j B_{s-j-1}B'_s M_{ba}\}] \\
\overset{0}{=}\ & -2E[tr\{\sum_{i>j}\sum_s \Delta A_s B'_{s-j-1}\psi'_i\psi_j B_{s-j-1}B'_s M_{ba}\}] \\
=\ & -2E[tr\{\sum_s \Delta A_s B'_s M_{ba}\}]\sum_{i>j}tr\{\psi'_i\psi_j\} \\
=\ & -2E[tr\{(M^+_{ab} - M_{ab})M_{ba}\}]tr\{\sum_{h=0}^{\infty}\gamma_\psi(h+1)\} \\
\overset{0}{=}\ & -(v_\psi - \tau_\psi)(tr\{E[M_3]\} - tr\{E[M_1]\}).
\end{aligned}
$$

Adding the contributions we find, see (47),

$$
\begin{aligned}
E[K_{11}] \overset{0}{=}\ & -2tr\{E[M_5]\psi'\psi\} - m_1(v_\theta + \tau_\theta - v_\psi + \tau_\psi) \\
& -m_3(v_\theta - \tau_\theta + v_\psi - \tau_\psi).
\end{aligned}
\tag{48}
$$

K$_{12}$

We can replace $M_{yy.d}$ by M_{yy} and find from (39)

$$
\begin{aligned}
E[K_{12}] \overset{0}{=}\ & 2E[tr\{M_{ay}(M_{yy} - TI_{n_y})M_{yb}M^{-1}_{bb}M_{ba}\}] \\
\overset{0}{=}\ & 2T^{-1}tr\{E[(M_{yy} - TI_{n_y})M_{yb}]E[M_{ba}M_{ay}]\} \\
=\ & 2tr\{\sum_{k=0}^{\infty}(\gamma(k+1)'\psi_k + \psi_k tr\{\gamma(k+1)'\})E[M_{ba}M^+_{ab}\psi']\} \\
\overset{0}{=}\ & 2tr\{M_3 V\}.
\end{aligned}
\tag{49}
$$

K$_{13}$

Next we consider

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

33

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

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

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

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

because M_{bd} is asymptotically independent of $M_{ba} M_{ab}^+$, see Lemma 6. We therefore evaluate

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

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

$$
E[K_{13}] \overset{0}{=} -2n_d E[tr\{M_3 \psi' \psi\}]. \tag{50}
$$

K_{14}

$$
E[K_{14}] = 2T^{-1} tr\{M_{ay} M_{yb}(M_{bb.d} - T I_{n_b}) M_{ba}\}.
$$

This term has the same stochastic components as K_{11}, and is of lower order of magnitude, such that

$$
E[K_{14}] \overset{0}{=} 0. \tag{51}
$$

Collecting terms we find from $(48), (49)$, and $(50), (51)$ the expression for $E[K_1]$

K2

We have

$$
E[K_2] = E[tr\{M_{ab}(M_{by.d} M_{yy.d}^{-1} M_{yb.d}) M_{ba}\}],
$$

and since $M_{by.d} M_{yy.d}^{-1} M_{yb.d} \overset{0}{=} M_{by} M_{yy}^{-1} M_{yb} \overset{0}{=} T^{-1} M_{by} M_{yb}$, we get

$$
E[K_2] \overset{0}{=} tr\{E[M_{ba} M_{ab}] E[M_{by} M_{yb}]\} \overset{0}{=} n_y m_1,
$$

34

see Lemma 5.

K3

Again we can drop the conditioning on d_t and replace M_{yy} by TI_{n_y} and find

$$E[K_3] \stackrel{0}{=} tr\{E[M_{ab}M_{ba}M_{ay}M_{ya}]\},$$

which by (38) can be written as

$$
\begin{aligned}
E[K_3] &\stackrel{0}{=} tr\{E[M_{ab}M_{ba}(M_{au}^+\theta' + M_{ab}^+\psi')(\theta M_{ua}^+ + \psi M_{ba}^+)]\} \\
&= tr\{\theta'\theta\}tr\{E[M_{ab}M_{ba}]\} + tr\{E[\psi'\psi M_{ba}^+ M_{ab}M_{ba}M_{ab}^+]\} \\
&= tr\{E[M_1]\}tr\{\theta'\theta\} + E[tr\{M_4\psi'\psi\}] = tr\{E[M_1]\}v_\theta + E[tr\{M_4\psi'\psi\}].
\end{aligned}
$$

K4

We drop the conditioning on d_t and replace M_{yy} by TI_{n_y} and find from Lemma 5

$$
\begin{aligned}
E[K_4] &\stackrel{0}{=} T^{-1}tr\{E[M_{ay}M_{yb}M_{by}M_{ya}]\} \\
&\stackrel{0}{=} T^{-1}tr\{E[(\theta M_{ua}^+ + \psi M_{ba}^+)(M_{au}^+\theta' + M_{ab}^+\psi')]E[M_{yb}M_{by}]\} \\
&= n_b(n_a tr\{\theta\theta'\} + E[tr\{M_6\psi'\psi\}]) = n_b(n_a v_\theta + E[tr\{M_6\psi'\psi\}]).
\end{aligned}
$$

This concludes the proof of Theorem 1. ■