A Bartlett correction factor for tests on the cointegrating relations

Johansen, Søren

Published in:
Econometric Theory

Publication date:
2000

Document version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Likelihood ratio tests for restrictions on cointegrating vectors are asymptotically $\chi^2$ distributed. For some values of the parameters this asymptotic distribution does not give a good approximation to the finite sample distribution. In this paper we derive the Bartlett correction factor for the likelihood ratio test and show by some simulation experiments that it can be a useful tool for making inference.

1. INTRODUCTION

In this paper we derive a Bartlett correction for tests on cointegrating relations in the vector autoregressive model for the $n$-dimensional process $X_t$ given by

$$\Delta X_t = \alpha (\beta' X_{t-1} + \rho' D_t) + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi d_t + \varepsilon_t, \quad t = 1, \ldots, T, \tag{1}$$

where $\varepsilon_t$ are independent and identically distributed (i.i.d.) $N_p(0, \Omega)$, the initial values are fixed, and $d_t$ ($n_d \times 1$) and $D_t$ ($n_D \times 1$) are deterministic terms, such as constant, linear term, etc. The matrices $\alpha$ and $\beta$ are $(n \times r)$ and the matrices $\Gamma_1, \ldots, \Gamma_{k-1}$, are $(n \times n)$, $\Phi (n \times n_d)$, and $\rho (n_D \times r)$. For the calculation of the expectation in this paper we assume that $X_t$ is $I(1)$. Conditions for this in terms of the parameters are given in Johansen (1996, Theorem 4.2).

The likelihood ratio test for hypotheses on $\beta$ has been treated in Johansen and Juselius (1990) and Ahn and Reinsel (1990), and it is known that it is asymptotically $\chi^2$ distributed, despite the fact that the asymptotic distribution of the estimator is mixed Gaussian. Finite sample distributions, however, are not always well approximated by asymptotic distributions (see, e.g., to mention a few of many studies of finite sample properties of tests of restrictions on $\beta$, Fachin, 1997; Gredenhoff and Jacobson, 1998; Jacobson, Vredin, and Warne, 1998; and Haug, 1998).

We derive here a correction term to the likelihood ratio test statistic for hypotheses on $\beta$ with the purpose of improving the approximation to the asymptotic $\chi^2$ distribution. The correction is the so-called Bartlett correction (see
Bartlett, 1937). For a recent survey of the theory of this type of correction see Cribari-Neto and Cordeiro (1996). Briefly, the method consists of calculating the expectation of the likelihood ratio (LR) statistic in the form $-2 \log LR$ for a given parameter point $\theta$ under the null hypothesis. Usually it is not possible to do this explicitly, and one can instead find an approximation of the form

$$E_{\theta}[-2 \log LR] = \frac{1}{f} \left( 1 + \frac{B(\theta)}{T} \right) + O(T^{-3/2}),$$

where $f$ is equal to the degrees of freedom for the test and $B(\theta)$ shows how the remaining parameters under the null hypothesis distort the mean and hence the distribution of the test statistic. The idea is that the quantity

$$\frac{-2 \log LR}{1 + \frac{B(\theta)}{T}}$$

has expectation $f + O_p(T^{-3/2})$, thus presumably a distribution that is closer to the limit distribution. Note that estimation of $\theta$ gives an extra error that is absorbed into the term $O_p(T^{-3/2})$. Lawley (1956) proved that under assumptions of i.i.d. variables, the same correction improves not only the mean but all moments, thus giving a mathematical explanation of why the correction works so well in practice. No similar theorem has been proved in the case of $I(1)$ variables. Still it is of interest to calculate the Bartlett correction to see how it works in practice.

The model (1) is characterized by dimension $(n)$, cointegrating rank $(r)$, lag length $(k)$, the number of deterministic terms restricted to the cointegrating space $(n_D)$, the number of unrestricted terms $(n_d)$ and finally of course the value of all the parameters and the sample size $(T)$.

The main result presented in Section 4 is that the Bartlett correction is a function of the parameters through only two functions, and various combinations of the preceding numbers. We find, for instance, for the test for the same restrictions on all cointegrating relations $\beta$, that is, $\beta = Hr(H(n \times s))$, that $f = r(n - s)$ and with $m = n + s - r + 1 + 2n_D$ we get

$$B(\theta) = \left[ \frac{1}{2} m + n_d + kn \right] + \frac{1}{r} [n - 2(r + 1) + m] v(\alpha) + 2(c(\alpha) + c_d(\alpha))].$$

The coefficients $v(\alpha), c(\alpha)$, and $c_d(\alpha)$ are given in Theorem 4, which follows.

It will be apparent from the examples and the simulations in Section 5 that the influence of the parameters is crucial. For some parameter values the usual $\chi^2$ approximation works well, whereas for others the correction factor is a useful improvement. Finally there are parameters points close to the boundary where the order of integration or the number of cointegrating relations change, and where the correction does not work well.
The plan of the paper is first to establish in Section 2 that a number of hypotheses can be given a general formulation as tests in a reduced rank regression model. In Section 3 an expansion is given of the estimators of this reduced rank regression. In Section 4 the main results on an expansion of the log likelihood ratio test and the Bartlett correction are given and the results specialized to the models discussed in Section 2, and finally in Section 5 some simulation experiments are conducted that show that the Bartlett correction is a useful addition to the usual asymptotic analysis. The very long and tedious proofs are given in an Appendix.

2. THE MODELS AND THE HYPOTHESES

We define in this section three models by restrictions on the cointegrating relations. All models can be analyzed by reduced rank regression (for a detailed analysis, see Johansen, 1996). The models allow deterministic terms of a suitably simple type, covering many usual situations. We show how correction terms for the tests of each of the models can be calculated if we have the correction term for a simple hypothesis. We show for each of the models how to formulate the test of a simple hypothesis as a test in a reduced rank regression, such that all the tests can be given the same uniform formulation. We first define the three models.

$\mathcal{M}_0$ Unrestricted Cointegrating Space. The model is given by equation (1) with unrestricted parameters.

$\mathcal{M}_1$ Same Restriction on All Cointegrating Relations. The model is defined as a submodel of $\mathcal{M}_0$ by the same restrictions on all cointegrating relations. This is expressed as

$$\beta = H\tau,$$

where $H$ is $(n \times s)$ of rank $s$ and known, $r \leq s < n$, and $\tau$ is $(s \times r)$ and unknown. The likelihood ratio test of $\mathcal{M}_1$ in $\mathcal{M}_0$ satisfies

$$-2 \log LR(\mathcal{M}_1|\mathcal{M}_0) \xrightarrow{w} \chi^2(r(n-s)).$$

The restrictions on $\beta$ can also be expressed as restrictions on $(\beta, \rho)$ in the form

$$\begin{pmatrix} \beta \\ \rho \end{pmatrix} = \begin{pmatrix} H & 0_{n \times r} \\ 0_{r \times s} & I_{n_{1D}} \end{pmatrix} \tau,$$

with $\tau ((s + n_{1D}) \times r)$. One could also define a model by restricting simultaneously both $\beta$ and $\rho$, but the present choice seems more relevant for the applications.
Some Cointegrating Relations Known. The model is defined by the restrictions
\[
\begin{pmatrix} \beta \\ \rho \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \rho_1 \end{pmatrix}, \begin{pmatrix} \beta_2^0 \\ \rho_2^0 \end{pmatrix},
\]
where the matrices \( \beta_2^0 \) \((n \times r_2)\) of rank \( r_2 \) and \( \rho_2^0 \) \((n_D \times r_2)\) are known and the matrices \( \beta_1 \) \((n \times r_1)\) and \( \rho_1 \) \((n_D \times r_1)\) are unknown \((r = r_1 + r_2)\), corresponding to prespecified coefficients \( \beta_2^0 \) and \( \rho_2^0 \) in some of the cointegrating relations. The likelihood ratio test of \( \mathcal{M}_2 \) in \( \mathcal{M}_0 \) satisfies
\[
-2 \log LR(\mathcal{M}_2 | \mathcal{M}_0) \xrightarrow{w} \chi^2(r_2(n + n_D - r)).
\]
The degrees of freedom is calculated in Johansen (1996, Theorem 7.3). It would also be relevant to formulate here the restriction that only \( \beta \) was partly known.

In the following sections we derive a correction factor for the test of a simple hypothesis on \( \beta \) and \( \rho \) in each of the models \( \mathcal{M}_0 \), \( \mathcal{M}_1 \), and \( \mathcal{M}_2 \) and apply these to derive a correction factor for the test of \( \mathcal{M}_1 \) in \( \mathcal{M}_0 \) and \( \mathcal{M}_2 \) in \( \mathcal{M}_0 \) using the following trick.

To test \( \mathcal{M}_1 \) in \( \mathcal{M}_0 \), say, we take parameters \( \beta^0 = H\tau^0 \) and \( \rho^0 \) corresponding to a parameter point in \( \mathcal{M}_1 \). We define the concentrated likelihood function \( L(\beta, \rho) \) and find the likelihood ratio test
\[
LR(\mathcal{M}_1 | \mathcal{M}_0) = \frac{\max_{\beta = H\tau, \rho} L(\beta, \rho)}{\max_{\beta, \rho} L(\beta, \rho)}
\]
\[
= \frac{\max_{\beta = \beta^0, \rho = \rho^0} L(\beta, \rho)}{\max_{\beta, \rho} L(\beta, \rho)} \frac{\max_{\beta = H\tau, \rho} L(\beta, \rho)}{\max_{\beta = H\tau, \rho} L(\beta, \rho)}
\]
\[
= LR(\beta = \beta^0, \rho = \rho^0 | \mathcal{M}_0)/LR(\beta = \beta^0, \rho = \rho^0 | \mathcal{M}_1),
\]
such that
\[
-2 \log LR(\mathcal{M}_1 | \mathcal{M}_0)
\]
\[
= -2 \log LR(\beta = \beta^0, \rho = \rho^0 | \mathcal{M}_0) + 2 \log LR(\beta = \beta^0, \rho = \rho^0 | \mathcal{M}_1).
\]
Hence we see that the correction for the test of \( \mathcal{M}_1 \) in \( \mathcal{M}_0 \) is the difference between the corrections to two tests of simple hypotheses on \( \beta \) and \( \rho \) in \( \mathcal{M}_0 \) and \( \mathcal{M}_1 \).
2.1. The Deterministic Terms

The correction will depend on the deterministic terms, and to get reasonably simple expressions we assume that they satisfy the relation

\[ d_{t+h} = M^h d_t, \quad h = \ldots, -1, 0, 1, \ldots \] (2)

for some matrix \( M \) with the property that

\[ |\text{eig}(M)| = 1. \] (3)

Further we assume that

\[ \Delta D_t = K' d_t \] (4)

for some \((n_d \times n_D)\) matrix \( K \). Finally we assume that \((D_t, d_t)_{t=1}^T\) are linearly independent. Thus we allow, for instance, \( d'_t = (1, t, t^2) \) and \( D_t = t^3 \), in which case

\[
M = \begin{pmatrix}
1 & 0 & 0 \\
1 & 1 & 0 \\
1 & 2 & 1
\end{pmatrix}, \quad K = \begin{pmatrix}
1 \\
-3 \\
3
\end{pmatrix}
\]

and \( M \) has eigenvalues equal to 1. If \( s_1, s_2, \) and \( s_3 \) are quarterly dummies we can consider combinations such as \( d'_t = (1, t, s_1(t), s_2(t), s_3(t)) \). In this case we have \( s_1(t + 1) = s_4(t) = 1 - s_1(t) - s_2(t) - s_3(t) \), \( s_2(t + 1) = s_1(t) \), and \( s_3(t + 1) = s_2(t) \) such that

\[
M = \begin{pmatrix}
1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 \\
1 & 0 & -1 & -1 & -1 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0
\end{pmatrix},
\]

which has eigenvalues \( \pm 1, \pm i \). Note that intervention dummies are not covered by this formulation and will give rise to more complicated formulae. The formulation also allows, with minor modifications, the possibility that \( n_D \) and \( n_d \) are zero.

LEMMA 1. If \( X_t \) is I(1) and given by equation (1) and if (2), (3), and (4) hold, then \( E(\beta' X_{t-1} + \rho' D_t) \) and \( E(\Delta X_t) \) are linear functions of \( d_t \).

Proof. From Granger’s representation theorem (see Johansen, 1996, Theorem 4.2), we find that because the process \( X_t \) is I(1) it has the representation

\[ X_t = C \sum_{i=1}^{t} (\varepsilon_i + \Phi d_t) + C(L)(\varepsilon_t + \alpha p' D_t + \Phi d_t) + A, \]

where \( \varepsilon_t \) is a white noise process with a unit root, and \( \Phi \) is a matrix of coefficients.
where \( C(z) = \sum_{i=0}^{\infty} C_i z^i \), \( A \) depends on initial conditions, \( \beta' A = 0 \), and
\[
C = \beta_{\perp} (\alpha_{\perp}' \Gamma_{\perp} \beta_{\perp})^{-1} \alpha_{\perp}'.
\]

Note that the condition that the process is \( I(1) \) implies that \( |\alpha_{\perp}' \Gamma_{\perp} \beta_{\perp}| \neq 0 \), such that \( C \) is well defined. It follows that
\[
E(\Delta X_t) = C\Phi d_t + C(L)(\alpha \rho' K' + \Phi \Delta)d_t
\]
\[
= \left[ C\Phi + \sum_{i=0}^{\infty} C_i (\alpha \rho' K'M^{-i} + \Phi(M^{-i} - M^{-i-1})) \right] d_t = \tilde{K}'d_t,
\]
say. Taking expectations in (1) we find
\[
\tilde{K}'d_t = \alpha E(\beta'X_{t-1} + \rho'D_t) + \sum_{i=1}^{k-1} \Gamma_i \tilde{K}'M^{-i}d_t + \Phi d_t,
\]
which shows the result for \( E(\beta'X_{t-1} + \rho'D_t) \). Note that the result that \( M^h \) grows at most as a polynomial in \( h \) (see Lemma 9 in the Appendix), shows that the sums are convergent, because \( C_i \) are exponentially decreasing.

We next show how simple hypotheses on \( \beta \) and \( \rho \) in \( \mathcal{M}_0, \mathcal{M}_1, \) and \( \mathcal{M}_2 \) give rise to regression equations that can be given the same formulation. This allows us to derive all the results from one general reduced rank regression equation.

### 2.2. A Simple Hypothesis on \( \beta \) and \( \rho \) in \( \mathcal{M}_0 \)

The model equation is given by (1), and we consider the hypothesis \( \beta = \beta^0, \rho = \rho^0 \), such that under the null hypothesis
\[
\Delta X_t = \alpha(\beta^0 X_{t-1} + \rho^0 D_t) + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi d_t + \varepsilon_t,
\]
which is easily estimated by regression of \( \Delta X_t \) on \( \beta^0 X_{t-1} + \rho^0 D_t \), lagged differences, and \( d_t \).

It is convenient for the calculations to reparametrize the model by defining new parameters and regressors that involve the true value. In the rest of this section we therefore need a notation for the true value of the parameters and also for the parameters of the model. We also need a notation for the estimator under the null hypothesis and one for the estimator under the alternative. Thus, for instance, we let \( \alpha \) denote the parameter, \( \alpha^0 \) the true value of the parameter, for which we calculate the expectations, \( \tilde{\alpha} \) the reduced rank estimator in the model, and \( \hat{\alpha} \) the regression estimator under the null hypothesis.
We use the notation
\[ \Psi = (\Gamma_1, \ldots, \Gamma_{k-1}), \quad \Gamma = I_n - \sum_{i=1}^{k-1} \Gamma_i. \]

Note that
\[ (I_n - CT)\beta_\perp = (I_n - \beta_\perp (\alpha_\perp' \Gamma \beta_\perp)^{-1} \alpha_\perp' \Gamma) \beta_\perp = 0, \]
such that for \( \bar{\beta} = \beta (\beta' \beta)^{-1} \)
\[ (I_n - CT) = (I_n - CT)(\bar{\beta} \beta' + \beta_\perp \bar{\beta}_\perp') = (I_n - CT)\bar{\beta} \beta'. \]

We therefore decompose the process into stationary and nonstationary components using the true values of the parameters:
\[ X_t = (I_n - C^0 \Gamma^0)\bar{\beta}^0 \rho^0 X_t + C^0 \Gamma^0 X_t. \]

It follows that
\[ \beta' X_{t-1} = \beta'(I_n - C^0 \Gamma^0)\bar{\beta}^0 \beta^0 X_{t-1} + \beta' \beta^0 (\alpha_\perp^0 \Gamma^0 \beta_\perp^0)^{-1} \alpha_\perp^0 \Gamma^0 X_{t-1}. \]

We define new parameters \( \phi \) and \( \delta = (\delta_1', \delta_2')' \) by
\[ \phi' = \beta'(I_n - C^0 \Gamma^0)\bar{\beta}^0 \quad (r \times r), \]
\[ \delta_1' = \beta' \beta_\perp^0 \quad (r \times (n - r)), \]
\[ \delta_2' = \rho' - \beta'(I_n - C^0 \Gamma^0)\bar{\beta}^0 \rho^0 \quad (r \times n_D), \]
such that the old parameters in terms of the new are given by
\[ \beta' = \delta_1' (\alpha_\perp^0 \Gamma^0 \beta_\perp^0)^{-1} \alpha_\perp^0 \Gamma^0 + \phi' \beta^0, \quad \rho' = \delta_2' + \phi' \rho^0. \]

The hypothesis \( \beta = \beta^0, \rho = \rho^0 \) is expressed in the new parameters as \( \delta = 0 \) and \( \phi = I_r \). Equation (1) with the new parameters is
\[ \Delta X_t = \alpha \phi' (\beta^0 X_{t-1} + \rho^0 D_t) + \alpha (\delta_1' (\alpha_\perp^0 \Gamma^0 \beta_\perp^0)^{-1} \alpha_\perp^0 \Gamma^0 X_{t-1} + \delta_2' D_t) \]
\[ + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi d_t + \epsilon_t. \]

We absorb \( \phi (r \times r) \) into \( \alpha (n \times r) \) and adjust \( \delta \) accordingly. The hypothesis of interest is then \( \delta = 0 \).

In the reduced rank regression (6) we use the result that \( \beta^0 X_{t-1} + \rho^0 D_t \) and \( \Delta X_t \) have a mean that is linear in \( d_t \) (see Lemma 1), and that \( \Phi \) enters unrestrictedly, to replace the regressors \( \beta^0 X_{t-1} + \rho^0 D_t \) and the lagged differences with the stationary regressors
\[ V_{t-1} = \beta^0 X_{t-1} - E_0 (\beta^0 X_{t-1}), \]
\[ Z_{t-1} = (\Delta X_{t-1} - E_0 (\Delta X_{t-1}), \ldots, \Delta X_{t-k+1} - E_0 (\Delta X_{t-k+1})). \]
We also want to replace the regressor \((\alpha_l^0 \Gamma^0 \beta_l^0)^{-1} \alpha_l^0 \Gamma^0 X_{t-1}\) by something simpler without changing the statistical model and hence the test that \(\delta = 0\). We find by summing equation (1) that

\[
\alpha_l(X_t - X_0) = \alpha_l \sum_{i=1}^{k-1} \Gamma_i (X_{t-i} - X_{-i}) + \alpha_l \sum_{i=1}^{t} (\varepsilon_i + \Phi d_i).
\]

By subtracting \(\sum_{i=1}^{k-1} \alpha_l \Gamma_i X_i\) on both sides and replacing \(t\) by \(t - 1\) we get

\[
\alpha_l \Gamma X_{t-1} = \alpha_l' X_0 + \alpha_l' \sum_{i=1}^{k-1} \Gamma_i (X_{t-i-1} - X_{t-i}) + \alpha_l' \sum_{i=1}^{t-1} (\varepsilon_i + \Phi d_i).
\]

Because we are correcting for lagged differences in regression (6) we can replace \((\alpha_l^0 \Gamma^0 \beta_l^0)^{-1} \alpha_l^0 \Gamma^0 X_{t-1}\) and \(D_t\) by the nonstationary regressor given by the common trends extended by \(D_t\):

\[
A_{t-1} = \begin{pmatrix} A_0 + \alpha_l^0' \sum_{i=1}^{t-1} (\varepsilon_i + \Phi^0 d_i) \\ D_t \end{pmatrix},
\]

where \(A_0\) depends on initial conditions.

Equation (6) in the new variables and with suitably redefined parameters becomes

\[
\Delta X_t = \alpha V_{t-1}^{(n)} + \alpha \delta' A_{t-1}^{(n-r)} + \Psi Z_{t-1}^{(n-r+n_p)} + \Phi d_t^{(k-1)n} + \varepsilon_t^{(n)},
\]

where the dimensions are indicated below each variable. The estimators for the parameters \(\delta, \alpha, \Psi, \Phi,\) and \(\Omega\) can be found by reduced rank regression of \(\Delta X_t\) on \((V_{t-1}, A_{t-1})\) corrected for \(Z_{t-1}\) and \(d_t\). Under the hypothesis \(\delta = 0\) the parameters can be found by regression of \(\Delta X_t\) on \(V_{t-1}, Z_{t-1},\) and \(d_t\).

Later we choose \(A_{t-1}\) such that it is orthogonal to the deterministic term \(d_t\), which simplifies some notation. Note that if \(D_t = 0\) then, of course, we do not extend the process, and \(A_{t-1}\) is defined entirely in terms of the random walks and initial values. Note also that if \(d_t\) contains a constant then, when correcting for \(d_t\), the initial values disappear.

**2.3. A Simple Hypothesis on \(\beta\) and \(\rho\) in Model \(\mathcal{M}_1\)**

In model \(\mathcal{M}_1\) the cointegrating vectors \(\beta\) are restricted as \(\beta = H\tau(s \times r)\), and equation (1) is

\[
\Delta X_t = \alpha (\tau' H' X_{t-1} + \rho' D_t) + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi d_t + \varepsilon_t.
\]

We consider again the simple hypothesis \(\beta = H\tau^0, \rho = \rho^0\), corresponding to a point in \(\mathcal{M}_1\). We want to show that by introducing the true parameters \(\beta^0,\)
\( \rho^0, \ldots \) as before we can reformulate the equations to have the form (10), such that a test of a simple hypothesis is again a test that \( \delta = 0 \).

We decompose the process \( X_{t-1} \) using the true value of the parameters and find

\[
\beta'X_{t-1} = \tau'H'(I_n - C^0\Gamma^0) \tilde{\beta}^0 \bar{b}^0 X_{t-1} + \tau'H'\beta^0_\perp (\alpha^0_\perp \Gamma^0 \beta^0_\perp)^{-1} \alpha^0_\perp \Gamma^0 X_{t-1}.
\]

In this case \( \beta^0 = H\tau^0 \), which implies that \( \beta^0_\perp = (H_\perp, H(H'H)^{-1} \tau^0_\perp) \) and hence

\[
\beta'\beta^0_\perp = \tau'H'\beta^0_\perp = \tau'\tau^0_\perp (0_{(s-r)\times(n-s)}, I_{s-r}).
\]

We introduce the new parameters \( \phi \) and \( \delta = (\delta'_1, \delta'_2)' \) by

\[
\phi' = \tau'H'(I_n - C^0\Gamma^0) \tilde{\beta}^0 (r \times r), \quad \delta'_1 = \tau'\tau^0_\perp (r \times (s-r)), \quad \delta'_2 = \rho' - \tau'H'(I_n - C^0\Gamma^0) \tilde{\beta}^0 \rho^0 (r \times n_D),
\]

because then

\[
\tau'H'X_{t-1} + \rho'D_t = \phi' (\beta^0_\perp X_{t-1} + \rho^0_\perp D_t) + \delta'_1 (0_{(s-r)\times(n-s)}, I_{s-r}) (\alpha^0_\perp \Gamma^0 \beta^0_\perp)^{-1} \alpha^0_\perp \Gamma^0 X_{t-1} + \delta'_2 D_t.
\]

The hypothesis is formulated as \( \delta = 0 \) and \( \phi = I_r \). We let \( V_{t-1} \) and \( Z_{t-1} \) be defined by (7) and (8) and replace in this case the \( (s-r) \)-dimensional non-stationary regressor \( (0_{(s-r)\times(n-s)}, I_{s-r}) (\alpha^0_\perp \Gamma^0 \beta^0_\perp)^{-1} \alpha^0_\perp \Gamma^0 X_{t-1} \) with \( s-r \) linear combinations \( K_1 \) of the common trends extended by \( D_t \):

\[
A_{t-1} = \left( K_0 + K_1 \alpha^0_\perp \sum_{i=1}^{r-1} (e_i + \Phi^0 d_i) \right) D_t
\]

for some matrices \( K_0 \) \((s-r)\times 1)\) depending on initial conditions and \( K_1 \) \((s-r)\times (n-r)\). Equation (11) then becomes

\[
\Delta X_t = \alpha V_{t-1} + \alpha \delta' \quad A_{t-1} + \Psi' Z_{t-1} + \Phi' d_t + e_t,
\]

where \( \delta' \) is absorbed in \( \alpha \) and the remaining parameters are adjusted accordingly such that \( Z_{t-1} \) and \( V_{t-1} \) also have mean zero. The hypothesis of interest is \( \delta = 0 \), which corresponds to \( \rho = \rho^0 \), \( \beta = \beta^0 = H\tau^0 \).

This equation has the same structure as (10) except that the dimension of \( A_{t-1} \) is changed to \( s-r + n_D \).
2.4. A Simple Hypothesis on $\beta$ and $\rho$ in Model $\mathcal{M}_2$

We decompose $\alpha$ corresponding to $\beta$ into $\alpha = (\alpha_1, \alpha_2)$ and find that the relevant part of (1) is

$$
\alpha(\beta'X_{t-1} + \rho'D_t) = \alpha_1(\beta_1'X_{t-1} + \rho_1'D_t) + \alpha_2(\beta_2'0X_{t-1} + \rho_2'0D_t).
$$

A simple hypothesis on $\beta$ and $\rho$ is formulated as $\beta_1 = \beta_1^0$, $\rho_1 = \rho_1^0$. Without loss of generality we can choose $\beta_1$ orthogonal to $\beta_2^0$ and adjust $\rho_1$ accordingly.

In this case we include the regressor $(\beta_2'0X_{t-1} + \rho_2'0D_t)$ with the lagged differences $Z_{t-1}$ instead of with $V_{t-1}$. We then decompose the first component $\beta_1'X_{t-1} + \rho_1'D_t$ of the process as before:

$$
\beta_1'X_{t-1} + \rho_1'D_t = \beta_1'(I_n - C^0\Gamma^0)\beta_2^0\rho_2'0X_{t-1} + \beta_1'\beta_2^0(\alpha_1^0'\Gamma^0\beta_2^0)^{-1}\alpha_1^0'\Gamma^0X_{t-1} + \rho_1'D_t.
$$

Now define $\phi = (\phi_1, \phi_2)$ by

$$
\beta_1'(I_n - C^0\Gamma^0)\beta_2^0\rho_2'0X_{t-1} = (\phi_1', \phi_2') \begin{pmatrix} \beta_1'0X_{t-1} \\ \beta_2'0X_{t-1} \end{pmatrix} = \phi_1'\beta_1'0X_{t-1} + \phi_2'\beta_2'0X_{t-1},
$$

so that with $A_{t-1}$ defined by (9)

$$
\beta_1'X_{t-1} + \rho_1'D_t = \phi_1'(I_n - C^0\Gamma^0)\beta_2^0 + \phi_2'(\beta_2'0X_{t-1} + \rho_2'0D_t) + \delta'A_{t-1},
$$

where we have defined the new parameters $\phi = (\phi_1', \phi_2')$ and $\delta = (\delta_1', \delta_2')'$. With

$$(\phi_1', \phi_2') = \beta_1'(I_n - C^0\Gamma^0)\beta_2^0 (r_1 \times r_1, (r_1 \times r_2),$$

$$\delta_1' = \beta_1'\beta_2^0 (r_1 \times (n - r),$$

$$\delta_2' = \rho_1' - \phi_1'\rho_1^0 - \phi_2'\rho_2^0 (r_1 \times n_D).$$

Note that $\delta = 0$ implies that $\beta_1$ is proportional to $\beta_1^0$, because $\beta_1$ is assumed orthogonal to $\beta_2^0$. Hence $(\phi_1', \phi_2')' = (\beta_1'\beta_2^0, 0)$ such that the hypothesis of interest in the new parameters is $\delta = 0$ and $\phi_1 = I_{r_1}$. The equation becomes

$$
\Delta X_t = \alpha_1 V_{t-1} + \alpha_1 \delta' A_{t-1} + \Psi Z_{t-1} + \Phi d_t + \varepsilon_t,
$$

(14)

where

$$
V_{t-1} = \beta_1^00X_{t-1},
$$

(15)

$$
Z_{t-1} = (X_{t-1}'\beta_2^0, \Delta X_{t-1}', \ldots, \Delta X_{t-k+1}')',
$$

(16)

both corrected for their mean, and where again $\alpha_1$, $\Psi$, and $\Phi$ have been redefined to absorb $\phi_1$ and to accommodate the changed regressors. It is seen that equation (14) is of the form (10), with a changed definition of $V_{t-1}$ and $Z_{t-1}$, because the assumed stationary combinations $\beta_2^00X_{t-1} - E_0(\beta_2^00X_{t-1})$ are moved to the lagged differences. The hypothesis can be tested as $\delta = 0$. 

Thus in a general formulation that covers all the hypotheses we are inter-
ested in, we need to allow the dimensions of the variables entering the equation
to be different from those given in (10). But we still need to preserve the prop-
erties that under the null hypothesis the process $Y_{t-1} = (V_{t-1}', Z_{t-1}')'$ is a mean
zero stationary autoregressive process and that $V_{t-1}$ and $\delta' A_{t-1}$ have the factor
$\alpha$ (or $\alpha_1$) in front. All models (10), (13), and (14) have the property that they
are estimated by reduced rank regression and that under the null hypothesis,
$\delta = 0$, the models are estimated by simple regression.

3. A REDUCED RANK EQUATION AND SOME EXPANSIONS

To cover the different cases considered in Section 2, we discuss expansion and
Bartlett correction of the likelihood ratio test for the hypothesis $\delta = 0$ in the equation

$$
\Delta X_t = \xi V_{t-1} + \xi \delta' A_{t-1} + \Psi Z_{t-1} + \Phi \ d_t + \varepsilon_t,
$$

(17)

where $\varepsilon_t$ are i.i.d. $N_n(0, \Omega)$ and the parameters $(\xi, \delta, \Psi, \Phi, \Omega)$ vary freely.

This notation covers the different situations considered for suitable choices
of $\xi$ and the regressors $V_{t-1}, A_{t-1},$ and $Z_{t-1}$ and their dimensions (see Table 1).

In all cases $A_{t-1}$ is a function of $D_t$ and $\xi_1 \sum_{i=1}^{t-1} (\varepsilon_i + \Phi d_i)$, with $\xi = \alpha$ or $\alpha_1$.
The variables $V_{t-1}$ and $Z_{t-1}$ are, under the hypothesis $\delta = 0$, stationary with
mean zero.

Note that the stacked $(r + (k - 1)n)$-dimensional process $Y_t = (V_t', Z_t')'$ is
the same for all cases and contains $\beta' X_t$ and the lagged differences corrected
for their mean.

In the rest of the paper we refer to the true value, the one for which we
calculate the expectation, without the superscript because that simplifies the
notation. We introduce some notation for the many product moments that are

<table>
<thead>
<tr>
<th>Model</th>
<th>$n_v$</th>
<th>$n_a$</th>
<th>$n_z$</th>
<th>$n$</th>
<th>$\xi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathcal{M}_0$</td>
<td>$r$</td>
<td>$n - r + n_D$</td>
<td>$(k - 1)n$</td>
<td>$n$</td>
<td>$\alpha$</td>
</tr>
<tr>
<td>$\mathcal{M}_1$</td>
<td>$r$</td>
<td>$s - r + n_D$</td>
<td>$(k - 1)n$</td>
<td>$n$</td>
<td>$\alpha$</td>
</tr>
<tr>
<td>$\mathcal{M}_2$</td>
<td>$r_1$</td>
<td>$n - r + n_D$</td>
<td>$r_2 + (k - 1)n$</td>
<td>$n$</td>
<td>$\alpha_1$</td>
</tr>
</tbody>
</table>
needed in the analysis of the expectation, but first we introduce the normalized errors

\[ B_t = (\xi' \Omega \xi) ^{-1/2} \xi' \varepsilon_t, \quad U_t = (\xi' \Omega^{-1} \xi)^{-1/2} \xi' \Omega^{-1} \varepsilon_t \]  

and define the product moment matrices \( M \), for the variables \( \Delta X_t, B_t, \varepsilon_t, U_t, \) and \( d_t \) at time \( t \) but \( V_{t-1}, A_{t-1}, \) and \( Z_{t-1} \) lagged one period. Thus, for instance,

\[
\sum_{t=1}^{T} \begin{pmatrix} \Delta X_t \\ V_{t-1} \\ \varepsilon_t \end{pmatrix} \left( \begin{pmatrix} \Delta X_t' \\ V_{t-1}' \\ \varepsilon_t' \end{pmatrix} \right) = \begin{pmatrix} M_{00} & M_{0v} & M_{0e} \\ M_{v0} & M_{vv} & M_{ve} \\ M_{e0} & M_{ev} & M_{ee} \end{pmatrix}.
\]

We also use the notation for any three process \( A_{t-1}, U_t, \) and \( V_{t-1}, \) say,

\[ M_{w,a} = M_{wv} - M_{aa} M_{va}^{-1} M_{vw}, \]

and in particular we use a notation for the moment matrices corrected for the lagged differences \( Z_{t-1} \) and \( d_t \), because many results look a bit simpler:

\[ S_{wv} = M_{wz,d} = M_{wv} - M_{ud} M_{vd}^{-1} M_{dv} - M_{za,d} M_{zd,v}^{-1} M_{zd,v}. \]

These moment matrices are natural when the likelihood function is concentrated with respect to \( \Psi \) and \( \Phi \).

The maximum likelihood estimators based upon (17) will be denoted by \( \tilde{\delta}, \tilde{\xi}, \) and \( \tilde{\Omega} \). The first order conditions for the estimators in model (17) can be solved for each of the variables as

\[ \tilde{\xi} = (S_{0v} + S_{0a} \tilde{\delta})(S_{vv} + \tilde{\delta}' S_{av} + S_{va} \tilde{\delta} + \tilde{\delta}' S_{aa} \tilde{\delta})^{-1}, \]  

\[ \tilde{\Omega} = T^{-1}(S_{00} - \tilde{\xi}(S_{vv} + \tilde{\delta}' S_{av} + S_{va} \tilde{\delta} + \tilde{\delta}' S_{aa} \tilde{\delta})\tilde{\xi}'), \]  

\[ \tilde{\delta} = S_{aa}^{-1}(S_{a0} - S_{aw} \tilde{\xi}')(\tilde{\Omega}^{-1} \tilde{\xi}(\tilde{\Omega}^{-1} \tilde{\xi})^{-1}. \]

Note that the equations cannot be solved simultaneously, because the estimators are expressed in terms of each other.

Under the null hypothesis \( \delta = 0 \), the estimators are denoted \( \hat{\xi} \) and \( \hat{\Omega} \):

\[ \hat{\xi} = S_{0v} S_{vv}^{-1} = \xi + S_{ee} S_{vv}^{-1}, \]

\[ \hat{\Omega} = T^{-1}(S_{00} - S_{0v} S_{vv}^{-1} S_{v0}) = T^{-1} S_{ee,v}. \]

We next expand the estimators \( \tilde{\xi}, \tilde{\Omega}, \) and \( \tilde{\delta}, \) not around the parameter point \( (\xi, \Omega, 0) \) but around the estimator under the null \( (\hat{\xi}, \hat{\Omega}, 0) \).
THEOREM 2. The estimators $\hat{\xi}$, $\hat{\Omega}$, and $\hat{\delta}$ can be expanded around $\tilde{\xi}$, $\tilde{\Omega}$, and 0, respectively:

$$\hat{\xi} - \tilde{\xi} = [S_{ae,u} \hat{\delta} - \hat{\xi} \hat{\delta}' S_{aa} \hat{\delta} - \tilde{\xi} \hat{\delta}' S_{av}] S_{uv}^{-1} + O_P(T^{-2}),$$

(22)

$$(\hat{\Omega} - \tilde{\Omega}) = T^{-1} \hat{\Omega} Q_0 + T^{-2} \hat{\Omega} Q_1 + O_P(T^{-5/2}),$$

(23)

$$\hat{\Omega} Q_0 = S_{e.a,u} \hat{\delta} \hat{\xi}' + \hat{\xi} \hat{\delta}' S_{ae,u} - \tilde{\xi} \hat{\delta}' S_{aa} \tilde{\delta} \hat{\xi}',$$

$$\hat{\Omega} Q_1 = [S_{e.a,u} \hat{\delta} - \hat{\xi} \hat{\delta}' S_{av} - \tilde{\xi} \hat{\delta}' S_{aa} \tilde{\delta}] (T^{-1} S_{av})^{-1} + \tilde{\delta} S_{aa}^{-1} S_{e.e.v} \tilde{\Omega}^{-1} \hat{\xi} (\tilde{\xi}' \hat{\Omega}^{-1} \hat{\xi})^{-1} + O_P(T^{-1} S_{aa}^{-1/2}).$$

(24)

The proof of Theorem 2 is given in the Appendix. Notice that if $n_D = 0$, then $S_{aa} \in O_P(T^2)$, but if $D_t$ is in the model the order of magnitude depends on $D_t$. Thus, we use $S_{aa}$ to indicate an order of magnitude, such that, for example, $\tilde{\delta} \in O_P(S_{aa}^{-1/2})$.

We conclude this section by giving the likelihood ratio test for $\delta = 0$ in model (17). We find

$$LR^{2/T}(\delta = 0) = \frac{|\tilde{\Omega}|}{|\hat{\Omega}|} = \frac{|\tilde{\Omega} - (\hat{\Omega} - \tilde{\Omega})|}{|\hat{\Omega}|} = |I_n - \hat{\Omega}^{-1} (\hat{\Omega} - \tilde{\Omega})|$$

such that by expanding $-2 \log LR$ and applying (23) we get

$$-2 \log LR = -T \log |I_n - \hat{\Omega}^{-1} (\hat{\Omega} - \tilde{\Omega})|$$

$$= \frac{1}{2} T \text{tr}\{\tilde{\Omega}^{-1} (\hat{\Omega} - \tilde{\Omega})\} + \frac{T}{2} \text{tr}\{(\hat{\Omega}^{-1} (\hat{\Omega} - \tilde{\Omega}))^2\},$$

$$= \frac{1}{2} \text{tr}\{Q_0\} + T^{-1} (\text{tr}\{Q_0\} + \frac{1}{2} \text{tr}\{Q_0^2\}).$$

(25)

We have used the notation $\overset{d}{\sim}$ to indicate that we have kept terms of the order $T^{-d}$.

4. THE BARTLETT CORRECTION FACTOR

In this section we give the main result on the Bartlett correction. We first discuss briefly the idea of conditioning on the common trends. Then we calculate some coefficients that are needed in the formulation of the main result. Next we give a stochastic expansion of $-2 \log LR$, and finally we state the main result on the Bartlett correction. The result is specialized to the various situations covered by the general formulation as discussed in Section 2.

We choose first to calculate the expectation of the likelihood ratio test statistic conditioning on the process $\xi_t^i$. The argument for this is that it is easier to do so because many of the expressions derived involve ratios of quadratic forms.
and turn out to be possible to calculate, if we first condition on $\xi_t^\perp \varepsilon_i$. Another argument is that the asymptotic distribution of $\hat{\beta}$ is mixed Gaussian, where the mixing variable is just the limit of $\sum_{i=1}^t \alpha_i^t \varepsilon_i$, which are fixed when we condition on $\xi_t^\perp \varepsilon_i$. The end result is that the first order correction term in the conditional mean does not depend on the conditioning variable, such that what we find is also a correction term for the unconditional mean.

4.1. The Conditioning Variables

The processes $A_{t-1}$ and $B_t$ (see (9), (12), and (18)) are defined in terms of $D_t$, $\xi_t^\perp \sum_{i=1}^{t-1} \varepsilon_i$, and $\xi_t^\perp \varepsilon_i$. It is convenient to orthogonalize $A_{t-1}$ on the deterministic terms $d_t$ such that in the following $M_{a d} = 0$. Note that if $d_t$ contains a constant, then $A_{t-1}$ no longer depends on the initial values, but in any case the limit of $T^{-1/2}A_{[Tu]}$ does not involve the initial values.

When we do not condition on $\xi_t^\perp \varepsilon_i$ we have the following relations (see Chan and Wei, 1988):

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

(26)

$$
T^{-1} \sum_{i=1}^T B_i B_{i-k}' \xrightarrow{a.s.} I_{n_b}, \quad \text{if } k = 0, \quad \text{and } 0 \text{ if } k \neq 0,
$$

(27)

$$
T^{-1} \sum_{i=1}^T B_i (B_{i-k}' KB_{i-k})B_i' \xrightarrow{a.s.} \text{tr}\{K\}I_{n_b}, \quad k \neq 0
$$

(28)

for any $n_b \times n_b$ matrix $K$. Finally we have the weak limit

$$
M_{ba} M_{aa}^{-1} M_{ab} \xrightarrow{w} \int_0^1 (dW) F' \left( \int_0^1 FF' du \right)^{-1} \int_0^1 F(dW)',
$$

(29)

where the Brownian motion $W(u)$ is defined by

$$
T^{-1/2} \sum_{i=1}^{[Tu]} B_i = (\xi_t^\perp \Omega \xi_t^\perp)^{-1/2} \xi_t^\perp T^{-1/2} \sum_{i=1}^{[Tu]} \varepsilon_t \xrightarrow{w} W(u),
$$

of dimension $n_b = n - n_v$. The process $F$ is defined as a linear transformation of the limit of $A_{t-1}$ (see Johansen, 1996). If, for instance, $\xi = \alpha$, $d_t = 1$, and $D_t = 0$ then

$$
A_{t-1} = \alpha_{t-1}^\perp \left( \sum_{i=1}^{t-1} \varepsilon_i - T^{-1} \sum_{i=1}^{t-1} \sum_{i=1}^t \varepsilon_i \right) + \alpha_{t-1}^\perp \Phi(t - \tau).
$$

For $\tau = \alpha_{t-1}^\perp \Phi (\neq 0)$ we see that the univariate process $\tau' A_{t-1}$ is dominated by the linear term, whereas the $n - r - 1$–dimensional process $\tau_{t-1}^\perp A_{t-1}$ has only
the random walk components. In this case \( F \) is of dimension \( n_b = n_a = n - r \) and is given by

\[
F_i(u) = W_i(u) - \int_0^1 W_i(u) du, \quad i = 1, \ldots, n_a - 1, \quad F_{n_a}(u) = u - \frac{1}{2}.
\]

When conditioning on the sequence \( \xi_{\perp} e_i \) we assume that relations (26)–(28) hold for the sequence we are fixing. We keep the notation \( M_{ba} M_{aa}^{-1} M_{ab} \) because the terms involving this quantity cancel in the final result.

### 4.2. The Autoregressive Model

Before we formulate the main result we need some notation for the vector autoregressive process given in model (1), which is the basis for all the calculations. Under the null hypothesis the model is estimated by ordinary least squares of \( D X_t \) on \( V_t, Z_t, d_t, \) and we therefore introduce the stacked process

\[
Y_t = PY_{t-1} + Q e_t,
\]

where

\[
P = \begin{pmatrix}
I_r + \beta' \alpha & \beta' \Gamma_1 & \ldots & \beta' \Gamma_{k-2} & \beta' \Gamma_{k-1} \\
0 & I_n & \ldots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \ldots & I_n & 0 \\
0 & 0 & \ldots & 0 & I_n
\end{pmatrix}, \quad Q = \begin{pmatrix}
\beta' \\
I_n \\
\vdots \\
0
\end{pmatrix}.
\]

(30)

We find the representation

\[
Y_t = \sum_{\nu = 0}^{\infty} P^\nu Q e_{t-\nu} = \sum_{\nu = 0}^{\infty} (\theta_{\nu} U_{t-\nu} + \psi_{\nu} B_{t-\nu}),
\]

where we have decomposed \( e_t \) into the components \( U_t \) and \( B_t \) (see (18)) and defined

\[
\theta_{\nu} = P^\nu Q \xi (\xi' \Omega^{-1} \xi)^{-1/2}, \quad \psi_{\nu} = P^\nu Q \Omega \xi_{\perp} (\xi_{\perp}' \Omega \xi_{\perp})^{-1/2}.
\]

We define the variance

\[
\Sigma = \text{Var}(Y_t) = \sum_{\nu = 0}^{\infty} (\theta_{\nu} \theta_{\nu}' + \psi_{\nu} \psi_{\nu}') = \Sigma_{\theta \theta} + \Sigma_{\psi \psi},
\]

(31)
say, and find

$$\theta = \sum_{\nu=0}^{\infty} \theta_{\nu} = (I_{n_{\nu}} - P)^{-1} Q\xi (\xi' \Omega^{-1} \xi)^{-1/2}, \quad (32)$$

$$\psi = \sum_{\nu=0}^{\infty} \psi_{\nu} = (I_{n_{\nu}} - P)^{-1} Q\Omega \xi_{\perp} (\xi_{\perp}' \Omega \xi_{\perp})^{-1/2}. \quad (33)$$

From the definition of $P$ and $Q$ it follows that

$$(I_{n_{\nu}} - P)(I_{r}, 0_{r \times (k-1)n})' = -Q\alpha.$$ 

By multiplying by the matrix $(I_{r}, 0_{r \times r_{2}})'$ we find because $\alpha_{1} = \alpha(I_{r}, 0_{r \times r_{2}})'$

$$(I_{n_{\nu}} - P)(I_{r}, 0_{r \times (r_{2}+(k-1)n)})' = -Q\alpha_{1},$$

such that with $\xi$ either $\alpha$ or $\alpha_{1}$ we have

$$(I_{n_{\nu}} - P)(I_{n_{\nu}}, 0_{n_{\nu} \times n_{\nu}})' = -Q\xi.$$ 

From (32) we then get with $\kappa_{\xi} = (\xi' \Omega^{-1} \xi)^{-1/2}$ that

$$\theta = -(I_{n_{\nu}}, 0_{n_{\nu} \times n_{\nu}})' \kappa_{\xi} = -\tilde{\kappa}_{\xi}, \quad (34)$$

say. Note that because $Y_{t} = (V_{t}', Z_{t}')'$ we get

$$\tilde{\kappa}_{\xi} \Sigma^{-1} \tilde{\kappa}_{\xi} = \kappa_{\xi} (I_{n_{\nu}}, 0_{n_{\nu} \times n_{\nu}}) \Sigma^{-1} (I_{n_{\nu}}, 0_{n_{\nu} \times n_{\nu}})' \kappa_{\xi} = \kappa_{\xi} \Sigma_{uv, v} \kappa_{\xi}. \quad (35)$$

4.3. The Main Results

We can finally state the main results. We start in Theorem 3 with a stochastic expansion of the likelihood ratio test for $\delta = 0$, which forms the basis for the calculation of the Bartlett correction factor in Theorem 4. The proofs are left for the Appendix, but we give here some corollaries to show explicitly how the correction can be used for the tests mentioned in Section 2. We first give the assumptions that are needed for the results to hold.

Assumption 1. We assume throughout that the process $X_{t}$ given by (1) is $I(1)$, that the deterministic terms satisfy (2)–(4), and that the conditioning variables $\xi_{t}' e_{t}$ and $\xi_{t}' \Sigma_{t-1} e_{t}$ satisfy (26)–(28).

THEOREM 3. Under Assumption 1, an expansion of the log likelihood ratio test for $\delta = 0$ based upon (17) is given by
\[-2 \log LR(\delta = 0) \]

\[
\frac{1}{2T} \text{tr}\{S_{auh,v}^{-1}S_{auh,v}S_{auh,v}^{-1}\} + \frac{1}{2T} \text{tr}\{(S_{uu}S_{ua}^{-1}S_{au})^2\} \\
+ 2T \text{tr}\{S_{aa}^{-1}S_{auv,b}S_{aa}^{-1}S_{vb}S_{bb}^{-1}S_{ba,v}\} \\
+ \text{tr}\{S_{aa}^{-1}S_{ab}^{-1}S_{aa}S_{ab}^{-1}S_{ba}^{-1}S_{aa}\} \\
+ \text{tr}\{S_{aa}^{-1}S_{ab}^{-1}S_{ba}^{-1}S_{aa}\} \\
- \text{tr}\{S_{aa}^{-1}S_{ab}^{-1}S_{ba}^{-1}S_{aa}\} \\
- 2 \text{tr}\{S_{aa}^{-1}S_{ab}^{-1}S_{ba}^{-1}S_{aa}\}.
\]

The first two terms are the test statistic for \(d = 0\) if \(\xi\) were known (see Johansen, 1999), the next term is of the order \(O_p(T^{-1/2})\) but has expectation \(O_p(T^{-1})\), and the last four terms are of the order \(O_p(T^{-1})\). The proof of Theorem 3, based upon the expansions in Theorem 2, is given in the Appendix.

**THEOREM 4.** Under Assumption 1, the conditional expectation given \(\xi'\perp\epsilon_t\) of the log likelihood ratio test for the hypothesis \(\delta = 0\) in (17) is approximated as

\[
E[-2 \log LR(\delta = 0)|\xi'\perp\epsilon_t] = n_vn_a \left[ \frac{1}{2T} \right] (n_v + n_a + 1) + n_a + n_a + n_v \\
+ \frac{n_a}{T} [(n - n_v + n_a - 1)v(\xi) + 2(c(\xi) + c_d(\xi))]
\]

where

\[
v(\xi) = \text{tr}\{V_\xi\}, \quad V_\xi = \tilde{\kappa}_\xi \kappa'_\xi \Sigma^{-1},
\]

\[
c(\xi) = \text{tr}\{P(I_{n_v} + P)^{-1}V_\xi\} + \text{tr}\{[P \otimes (I_{n_v} - P)V_\xi][I_{n_a} \otimes I_{n_a} - P \otimes P]^{-1}\},
\]

\[
c_d(\xi) = \text{tr}\{[M \otimes (I_{n_v} - P)V_\xi][I_{n_a} \otimes I_{n_a} - M \otimes P]^{-1}\}.
\]

Here \(\tilde{\kappa}_\xi\) is given by (34), \(P\) is given by (30), and \(M\) is defined in Section 2.1.

Note that the conditional expectation does not depend on the conditioning variable and hence equals the expectation. It will be seen from the proof that the correction term is the one derived in the situation where \(\xi\) is known (see Johansen, 1999), apart from a term equal to \(T^{-1}n_a(n - n_v)v(\xi)\). Note that the coefficients \(v, c,\) and \(c_d\) depend on the choice of \(\xi\). If \(\xi = \alpha\), then from (35)

\[
v(\alpha) = \text{tr}\{(\alpha'\Omega^{-1}\alpha)^{-1}\Sigma^{-1}\}.
\]
with $\Sigma_{\beta \beta} = \text{Var}(\beta'X_t|\Delta X_t,\ldots,\Delta X_{t-k+2})$. If, however, $\xi = \alpha_1$ then

$$v(\alpha_1) = \text{tr}\{[\alpha_1'\Omega^{-1}\alpha_1]^{-1}\Sigma_{\beta_1\beta_1}^{-1}\},$$

with $\Sigma_{\beta_1\beta_1} = \text{Var}(\beta'_1X_t|\beta'_2X_t,\Delta X_t,\ldots,\Delta X_{t-k+2})$, corresponding to having moved $\beta'_2X_{t-1}$ from $V_{t-1}$ to $Z_{t-1}$.

The coefficient $c_d(\xi)$ can be calculated simply in some cases, such as $d_t = (1,t)'$, because then $\text{tr}\{M^h\} = n_d = 2$ for all $h$. This means that

$$c_d(\xi) = \text{tr}\{[M \otimes (I_{n_y} - P)V_{\xi}][I_{n_y} \otimes I_{n_y} - M \otimes P]^{-1}\}$$

$$= \sum_{v=0}^{\infty} \text{tr}\{[M \otimes (I_{n_y} - P)V_{\xi}][M^v \otimes P^v]\}$$

$$= \sum_{v=0}^{\infty} \text{tr}\{M^{v+1}\} \text{tr}\{(I_{n_y} - P)V_{\xi}P^v\} = n_d \text{tr}\{V_{\xi}\} = n_d v(\xi).$$

Note that $M$ can be replaced by $M'$ because only $\text{tr}\{M^{v+1}\}$ enters the result. If $d_t$ contains seasonal dummies then $\text{tr}\{M^h\}$ is a periodic function and a more complicated expression can be found. To understand the parameter function $V_{\xi}$ that enters the expressions, note that the long-run variance of $Y_t$ conditional on the common trends is given by $\theta' = \tilde{\kappa}_\xi \tilde{\kappa}_\xi'$ (see (34)). Thus the matrix $V_{\xi} = \tilde{\kappa}_\xi \tilde{\kappa}_\xi'\Sigma^{-1}$ measures the “ratio” between the conditional long-run variance and the unconditional variance of $Y_t$.

We specialize the result to the hypotheses discussed in Section 2.

**COROLLARY 5.** The Bartlett correction for the test of a simple hypothesis $\beta = \beta^0, \rho = \rho^0$ in model $\mathcal{M}_0$, is given by

$$E[-2 \log LR(\beta = \beta^0, \rho = \rho^0|\mathcal{M}_0)]$$

$$= \frac{1}{2} r(n - r + n_D) + \frac{r(n - r + n_D)}{T}\left[\frac{1}{2}(n + n_D + 1) + n_d + kn\right]$$

$$+ \frac{(n - r + n_D)}{T}\left[(2(n - r) + n_D - 1)v(\alpha) + 2(c(\alpha) + c_d(\alpha))\right]$$

where $v(\alpha), c(\alpha),$ and $c_d(\alpha)$ are given in Theorem 4.

Proof. This follows from Theorem 4 by substituting $n_{u} = r, n_{a} = n - r + n_D, n_{z} = (k - 1)n, \xi = \alpha$ (see Table 1). \[\blacksquare\]

**COROLLARY 6.** The Bartlett correction factor for the test of $\mathcal{M}_1: \beta = H\tau$, with $H(n \times s)$, is given by
\[ E[-2 \log LR(M_1 | M_0)]/r(n - s) \]
\[ \leq 1 + \frac{1}{T} \left[ \frac{1}{2} (n + s - r + 1 + 2n_D) + n_d + kn \right] \]
\[ + \frac{1}{Tr} [(2n + s - 3r - 1 + 2n_D)v(\alpha) + 2(c(\alpha) + c_d(\alpha))], \]

where \( v(\alpha), c(\alpha), \text{ and } c_d(\alpha) \) are given in Theorem 4.

**COROLLARY 7.** The Bartlett correction factor for the test of \( M_2, \)
\[
\left( \begin{array}{c} \hat{\beta} \\ \rho \end{array} \right) = \left( \left( \begin{array}{c} \beta_1 \\ \rho_1 \end{array} \right), \left( \begin{array}{c} \beta_2^0 \\ \rho_2^0 \end{array} \right) \right),
\]

where the matrix \( \beta_2^0 (n \times r_2) \) is of rank \( r_2 \) and \( \rho_2^0 \) is \( (1 \times r_2) \), is given by

\[ E[-2 \log LR(M_2 | M_0)]/r_2(n - r + n_D) \]
\[ \leq 1 + \frac{1}{T} \left[ \frac{1}{2} (n + n_D + 1 - r_1) + n_d + kn \right] \]
\[ + \frac{1}{Tr_2} [(2(n - r) + n_D - 1)(v(\alpha) - v(\alpha_1)) - r_2v(\alpha_1)] \]
\[ + \frac{2}{Tr_2} [(c(\alpha) - c(\alpha_1) + c_d(\alpha) - c_d(\alpha_1))], \]

where the coefficients \( c(\cdot), c_d(\cdot), \text{ and } v(\cdot) \) are defined in Theorem 4.

The proofs of these corollaries follow from Theorem 4 by using the trick explained in Section 2.

**5. SIMULATION EXPERIMENTS**

We report here some simple simulation experiments to illustrate the usefulness of the correction. We first give the result for the model with only one lag and one cointegrating relation, because we get complete information on how the correction works. We then present a few results where the data generating process (DGP) has been chosen so as to match the results obtained for real data analyzed elsewhere. Throughout the number of simulations is 10,000. When calculating the Bartlett correction we assume that we know the lag length and the cointegrating rank as given by the DGP, but we estimate the remaining parameters and hence the Bartlett factor.
5.1. The Model with One Cointegrating Vector and One Lag

The model is

\[ \Delta X_t = \alpha \beta' X_{t-1} + \varepsilon_t. \]

We find the coefficients

\[ \text{Var}(\beta' X_t) = \Sigma = \frac{\beta' \Omega \beta}{1 - (1 + \beta' \alpha)^2}, \quad \kappa_\alpha^2 = (\alpha' \Omega^{-1} \alpha)^{-1}, \]

\[ v(\alpha) = V_{\alpha} = \Sigma^{-1} \kappa_\alpha^2 = -\frac{\beta' \alpha (2 + \beta' \alpha)}{\alpha' \Omega^{-1} \alpha \beta' \Omega \beta}, \]

\[ P = 1 + \beta' \alpha, \quad c(\alpha) = -2 \frac{\beta' \alpha (1 + \beta' \alpha)}{\alpha' \Omega^{-1} \alpha \beta' \Omega \beta}. \]

With this notation we find the following corollary from Corollary 5.

**COROLLARY 8.** In the model \( \Delta X_{t-1} = \alpha \beta' X_{t-1} + \varepsilon_t \) with one cointegrating relation, the Bartlett correction factor for the hypothesis \( \beta = \beta^0 \) is

\[
E[-2 \log LR(\beta = \beta^0 | \mathcal{M}_0)] / (n - 1)
\]

\[
= 1 + \frac{1}{2T} (3n + 1) - \frac{1}{T} \frac{\beta' \alpha}{\alpha' \Omega^{-1} \alpha \beta' \Omega \beta}
\times [(2n - 3)(2 + \beta' \alpha) + 4(1 + \beta' \alpha)].
\]

It is seen that the parameters enter through two functions \( \alpha' \beta \) and \( \alpha' \Omega^{-1} \alpha \beta' \Omega \beta \). This result holds not only for the correction term but for the distribution of the likelihood ratio statistic. To see this we transform the problem linearly by defining \( \tilde{X}_t = v' X_t \), where \( v_1 = \beta (\beta' \Omega \beta)^{-1/2}, \)

\[ v_2 = -(\Omega^{-1} - \beta (\beta' \Omega \beta)^{-1} \beta') \alpha (\alpha' \Omega^{-1} \alpha - \alpha' \beta (\beta' \Omega \beta)^{-1} \beta' \alpha)^{-1/2}, \]

and finally \( v_3, \ldots, v_n \) are such that

\[
v'^\prime \Omega v_j = \begin{cases} 1 & i = j, \\ 0 & i \neq j, \end{cases} \quad i, j = 1, \ldots, n.
\]

The new variables satisfy the equations

\[
\Delta \tilde{X}_{1t} = \eta \tilde{X}_{1t-1} + \tilde{\varepsilon}_{1t},
\]

\[
\Delta \tilde{X}_{2t} = \gamma \tilde{X}_{1t-1} + \tilde{\varepsilon}_{2t},
\]

\[
\Delta \tilde{X}_{it} = \tilde{\varepsilon}_{it}, \quad i = 3, \ldots, n,
\]

(36)

where \( \tilde{\varepsilon}_t = v' \varepsilon_t \) are i.i.d. \( N_n(0, I_n) \) and

\[ \eta = \beta' \alpha, \quad \gamma = -(\alpha' \Omega^{-1} \alpha \beta' \Omega \beta - (\alpha' \beta)^2)^{1/2}, \]
such that

\[ v(\alpha) = -\frac{\eta(2 + \eta)}{\gamma^2 + \eta^2}, \quad c(\alpha) = -2 \frac{\eta(1 + \eta)}{\gamma^2 + \eta^2}. \]

Thus any DGP given by \((\alpha, \beta, \Omega)\) is completely described by just two numbers and can be transformed into the “canonical” form (36) with \(\beta' = (1,0,\ldots,0)\) and \(\alpha' = (\eta, \gamma, 0,\ldots,0)\). It is therefore possible for a given \(n\) to tabulate the effect of the Bartlett correction as a function of two variables \(\gamma\) and \(\eta\) (see Table 2, where we give some results for \(T = 50, n = 2,\) and \(n = 5\)). It is seen that for \(n = 2\) a nominal 5% test can have an actual size up to 16% and that in many cases (roughly \(\eta + \gamma < -0.2\)) the Bartlett correction factor gives a useful correction.

Note that for \(\gamma = 0\), both coefficients \(v(\alpha)\) and \(c(\alpha)\) have a factor \(\eta^{-1}\), which tends to infinity for small \(\eta\). The DGP where both \(\gamma\) and \(\eta\) are zero corresponds to no cointegration, and the test on \(\beta\) does not have a meaning in such a situation. The model with \(\eta = 0\), and \(\gamma \neq 0\), corresponds to a DGP generating an \(I(2)\) process, and the derivation of the correction factor is not valid in this case.

For \(n = 5\) the situation is worse, and the actual size can be very large indeed. The region where the Bartlett correction is useful is approximately given by \(\eta + \gamma < -0.4\). Obviously the situation improves if \(T\) increases.

Usually the test for \(\beta\) is preceded by a test for the rank, and if \(\eta\) and \(\gamma\) are sufficiently small the hypothesis of one cointegrating relation will not be accepted; thus for small values of \(\gamma\) and \(\eta\) the Bartlett correction is not needed.

5.2. Some Real Life Examples

In the examples that follow, taken from Johansen (1996), we use as the DGP the AR(2) model fitted to the real data. We have left out the seasonal dummies and started the process at the initial values for the data.

We first consider the Danish data consisting of the four variables \(m_t\) (log real M2), \(y_t\) (log real income), \(i^b_t\) (bond rate), and \(i^d_t\) (deposit rate) observed quarterly from 1974:1 to 1987:3. We fitted a model with a restricted constant term and simulated a time series with 53 observations, which was the number of observations in the example.

We first give the result for a simple test on \(\beta\), that is, that the coefficients are those of the DGP. The Bartlett factor in this case is given by Corollary 6, because we have a hypothesis that involves \(\beta\) and not \(\rho\). We find with \(n = 4, r = 1, s = 1, n_D = 1, n_d = 0, k = 2\) that the degrees of freedom is \(r(n - s) = 3\) and that

\[
E[-2 \log LR(\beta = \beta^0 | \mathcal{M}_0)]/3 = 1 + \frac{23}{2T} + \frac{1}{T}[7v(\alpha) + 2c(\alpha)].
\]
### Table 2. Corrected size over the uncorrected size for a nominal 5% test ($T = 50$, $r = 1$, and $n = 2$ and $5$)

<table>
<thead>
<tr>
<th>$\gamma \backslash \eta$</th>
<th>$n = 2$</th>
<th></th>
<th></th>
<th></th>
<th>$n = 5$</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$-0.1$</td>
<td>$-0.2$</td>
<td>$-0.4$</td>
<td>$-0.8$</td>
<td>$-0.1$</td>
<td>$-0.2$</td>
<td>$-0.4$</td>
<td>$-0.8$</td>
</tr>
<tr>
<td>0.0</td>
<td>1.3/16.5</td>
<td>3.7/12.2</td>
<td>4.7/8.6</td>
<td>5.2/6.8</td>
<td>0.02/78.6</td>
<td>1.8/64.6</td>
<td>6.8/40.3</td>
<td>6.2/17.6</td>
</tr>
<tr>
<td>$-0.1$</td>
<td>3.8/13.5</td>
<td>4.2/11.0</td>
<td>4.8/8.7</td>
<td>4.9/7.0</td>
<td>1.7/67.9</td>
<td>3.7/59.2</td>
<td>7.0/38.5</td>
<td>6.2/17.5</td>
</tr>
<tr>
<td>$-0.2$</td>
<td>5.4/10.8</td>
<td>4.9/9.9</td>
<td>4.9/8.6</td>
<td>5.1/7.0</td>
<td>9.5/44.9</td>
<td>7.4/46.8</td>
<td>7.5/34.1</td>
<td>6.1/17.2</td>
</tr>
<tr>
<td>$-0.4$</td>
<td>6.1/8.2</td>
<td>5.6/8.3</td>
<td>5.4/8.0</td>
<td>5.2/7.1</td>
<td>8.3/20.0</td>
<td>7.8/24.1</td>
<td>7.0/23.6</td>
<td>6.0/15.5</td>
</tr>
<tr>
<td>$-0.8$</td>
<td>5.6/6.9</td>
<td>5.5/7.2</td>
<td>5.8/7.4</td>
<td>5.4/6.9</td>
<td>6.1/11.5</td>
<td>6.0/13.2</td>
<td>5.8/13.9</td>
<td>5.6/12.8</td>
</tr>
</tbody>
</table>
We find for a test of nominal size 5% an actual size of 20.7% and a corrected size of 8.0%.

Another test of the form $\beta = H\tau$ is given by the matrix $H$:

$$H = \begin{pmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \end{pmatrix}$$

with $\tau$ ($2 \times 1$) corresponding to the test that $m_t$ and $y_t$ enter with the same coefficient with opposite sign and that the same holds for $i_t^b$ and $i_t^d$.

We find again from Corollary 6 with $n = 4, r = 1, s = 2, k = 2, n_D = 1, n_d = 0$ that the Bartlett factor is

$$E[-2 \log LR(\mathcal{M}_1|\mathcal{M}_0)]/2 \overset{!}{=} 1 + \frac{12}{T} + \frac{1}{T} [8v(\alpha) + 2c(\alpha)].$$

We find that a nominal 5% test gives an actual size of 19.5%, whereas the size for the corrected test is 8.9%.

As another example consider the data consisting of consumer price indices for Australia, $p_t^{au}$, and the United States, $p_t^{us}$, and the exchange rate $\text{exch}_t$ (all in logarithms) together with the five year treasury bond rate for both countries, $i_t^{au}$ and $i_t^{us}$. The data are observed quarterly from 1972:1 to 1991:1, which gives an effective number of observations of 75. We fitted a model with two lags and unrestricted constant and found two cointegrating relations. We first test a simple hypothesis on the two cointegrating relations. In this case we have $n = 5, r = 2, s = 2, k = 2, n_D = 0, n_d = 1$, such that the degrees of freedom is $r(n - s) = 6$. Because $d_t = 1$, we find that $c_d(\alpha) = n_d v(\alpha) = v(\alpha)$.

The Bartlett factor can be found from Corollary 6 and is given by

$$E[-2 \log LR(\mathcal{M}_1|\mathcal{M}_0)]/6 \overset{!}{=} 1 + \frac{14}{T} + \frac{1}{2T} [7v(\alpha) + 2c(\alpha)].$$

We found that a nominal 5% test in reality corresponds to a size of 21%. The correction of the test gives a size of 6.3%.

Next consider the test for price homogeneity given by the restriction $R'\beta = 0$, where

$$R = (1,1,0,0,0,0)$$

and $H = R_1$. In this example $s = 4$ such that the degrees of freedom is $r(n - s) = 2$. We find the Bartlett factor from Corollary 6 given by

$$E[-2 \log LR(\mathcal{M}_1|\mathcal{M}_0)]/2 \overset{!}{=} 1 + \frac{15}{T} + \frac{1}{2T} [9v(\alpha) + 2c(\alpha)].$$
A nominal 5% test corresponds to a test of size 10.5%, and the Bartlett correction gives the size 3.4%.

6. CONCLUSION

In this paper we have derived a correction to the log likelihood ratio statistic for various hypotheses on the cointegrating coefficients in a cointegrated vector autoregressive model. The correction depends on the parameters under the null hypothesis and on dimension, lag length, cointegrating rank, number of restricted deterministic terms, and number of unrestricted terms. The effect of the parameters is summarized in two functions, which are easily calculated once the parameters of the model have been estimated.

The usefulness of the results is demonstrated by some simulation experiments.

REFERENCES


APPENDIX

Some technical results. The deterministic terms are modelled by the matrix $M$, which enters in the proof of Lemma 1. We there need the following result.
LEMMA 9. Under the assumptions (2) and (3) the powers $M^h$ grow at most as a polynomial in $h$.

**Proof.** The matrix $M$ has a Jordan form that is block diagonal with blocks of the form

$$J_k(\lambda) = \lambda I_k + J, \quad J_{ij} = 1, \quad i = j + 1, \quad J_{ij} = 0, \quad i \neq j + 1,$$

where $\lambda$ is an eigenvalue of $M$, assumed to satisfy $|\lambda| = 1$. In the powers of $J$ the subdiagonal moves down, and $J^k = 0$. We find that the Jordan form for $M^h$ has the blocks

$$J_k(\lambda)^h = (\lambda I_k + J)^h = \sum_{i=0}^{k-1} \binom{h}{i} \lambda^{h-i} J^i.$$

This is bounded by a polynomial of degree $k - 1$ in $h$. This proves that $M^h$ is bounded by a polynomial of degree at most equal to the order (minus one) of the largest Jordan block in $M$. Because $M^{-1}$ has eigenvalues $\lambda$, the same result holds for negative powers of $M$. \(\blacksquare\)

When regressing a stationary variable on deterministic terms $d_t$, the order of magnitude of the product moments depends on $d_t$. We formulate a result that is used in the proof of Theorem 4.

LEMMA 10. Let $S_t = \sum_{i=0}^{\infty} \theta_i e_{t-i}$ with $\theta_i$ decreasing exponentially. Let

$$\gamma(h) = \text{Cov}(S_t, S_{t+h}) = \sum_{i=0}^{\infty} \theta_i \Omega^{i+h}.$$

Let $d_t$ satisfy $d_{t+1} = M d_t$, with $|\text{eig}(M)| = 1$; then

$$\text{tr}\{E(M_{sd} M_{dd}^{-1} M_{ds})\} \rightarrow \sum_{h=-\infty}^{\infty} \text{tr}\{M^h\} \text{tr}\{\gamma(h)\}. \quad (A.1)$$

**Proof.**

$$\text{tr}\{E(M_{sd} M_{dd}^{-1} M_{ds})\} = \text{tr}\left\{E \sum_{i=1}^{T} \sum_{i,j=0}^{\infty} \theta_i \theta_j \epsilon_{i-j} d_i^t M_{dd}^{-1} d_s \epsilon'_{s-j} \theta_j^t \right\}$$

$$= \sum_{i,j,t} \text{tr}\{\theta_i \theta_j^t\} \text{tr}\{d_i^t M_{dd}^{-1} d_{i-j}^{t+i} \}$$

$$= \sum_{i,j} \sum_{i=1}^{T} \text{tr}\{d_i^t M_{dd}^{-1} d_i M_{dd}^{-1} \text{tr}\{\theta_j^t \theta_i\}}$$

$$\rightarrow \sum_{h=-\infty}^{\infty} \text{tr}\{M^h\} \text{tr}\{\gamma(h)\}. \quad \blacksquare$$
Proof of Theorem 2 and some expansions of a projection matrix

Proof of Theorem 2.

Proof of (22). We expand (19) and find with $S_{0a} = \xi S_{va} + S_{ea}$, $S_{0v} = \xi S_{ev} + S_{ev}$, and $\hat{\xi} - \xi = S_{ev} S_{ev}^{-1}$ that

$$
\hat{\xi} = \hat{\xi} + [S_{0a} \hat{\delta} - \hat{\xi}(\hat{\delta}' S_{ev} + S_{va} \hat{\delta} + \hat{\delta}' S_{aa} \hat{\delta})] S_{ev}^{-1} + O_p(T^{-2})
$$

$$
= \hat{\xi} + [S_{ea} \hat{\delta} - \hat{\xi} \hat{\delta}' S_{aa} \hat{\delta} - \hat{\delta}' S_{av} S_{ev}^{-1} + O_p(T^{-2}),
$$

where we have used the identity

$$
S_{0a} - \hat{\xi} S_{va} = S_{ea} - (\hat{\xi} - \xi) S_{va} = S_{ea} - S_{ev} S_{ev}^{-1} S_{va} = S_{ea} \hat{\delta}.
$$

(A.2)

Proof of (23). From (19) and (20) we find

$$
T \hat{\Omega} = S_{00} - (S_{0v} + S_{0a} \hat{\delta})(S_{ev} + \hat{\delta}' S_{av} + S_{va} \hat{\delta} + \hat{\delta}' S_{aa} \hat{\delta})^{-1}(S_{v0} + \hat{\delta}' S_{a0}).
$$

We now expand the last term and keep terms of order $T^{-1}$. Throughout we use $\hat{\xi} = S_{0v} S_{ev}^{-1}$. Then

$$
(S_{0v} + S_{0a} \hat{\delta})(S_{ev} + \hat{\delta}' S_{av} + S_{va} \hat{\delta} + \hat{\delta}' S_{aa} \hat{\delta})^{-1}(S_{v0} + \hat{\delta}' S_{a0})
$$

$$
= (S_{0v} + S_{0a} \hat{\delta})[S_{ev}^{-1} - S_{ev}^{-1}(\hat{\delta}' S_{av} + S_{va} \hat{\delta} + \hat{\delta}' S_{aa} \hat{\delta}) S_{ev}^{-1}
$$

$$
+ S_{ev}^{-1}((\hat{\delta}' S_{av} + S_{va} \hat{\delta} + \hat{\delta}' S_{aa} \hat{\delta}) S_{ev}^{-1})^2](S_{v0} + \hat{\delta}' S_{a0})
$$

$$
= S_{0v} S_{ev}^{-1} S_{v0} + \hat{\xi} \hat{\delta}' S_{a0} + S_{0a} \hat{\delta} \hat{\delta}' - \hat{\xi}(\hat{\delta}' S_{av} + S_{va} \hat{\delta} + \hat{\delta}' S_{aa} \hat{\delta}) \hat{\delta}'
$$

$$
+ S_{0a} \hat{\delta} S_{ev}^{-1} \hat{\delta}' S_{a0} - \hat{\xi}(\hat{\delta}' S_{av} + S_{va} \hat{\delta} + \hat{\delta}' S_{aa} \hat{\delta}) S_{ev}^{-1} \hat{\delta}' S_{0a}
$$

$$
- S_{0a} \hat{\delta} S_{ev}^{-1} (\hat{\delta}' S_{av} + S_{va} \hat{\delta} + \hat{\delta}' S_{aa} \hat{\delta}) \hat{\delta}'
$$

$$
+ \hat{\xi}(\hat{\delta}' S_{av} + S_{va} \hat{\delta} + \hat{\delta}' S_{aa} \hat{\delta}) S_{ev}^{-1} (\hat{\delta}' S_{av} + S_{va} \hat{\delta} + \hat{\delta}' S_{aa} \hat{\delta}) \hat{\delta}'.
$$

The main term combines with $S_{00}$ to give $T \hat{\Omega} = S_{00} - S_{0v} S_{ev}^{-1} S_{v0}$. The term of order $T^0$ is

$$
\hat{\xi} \hat{\delta}' S_{a0} + S_{0a} \hat{\delta} \hat{\delta}' - \hat{\xi}(\hat{\delta}' S_{av} + S_{va} \hat{\delta} + \hat{\delta}' S_{aa} \hat{\delta}) \hat{\delta}' = \hat{\xi} \hat{\delta}' S_{ea} \hat{\delta}' + S_{ea} \hat{\delta} \hat{\delta}'.
$$

where we have used (A.2). The term of order $T^{-1}$ is

$$
[S_{0a} \hat{\delta} - \hat{\xi}(\hat{\delta}' S_{av} + S_{va} \hat{\delta} + \hat{\delta}' S_{aa} \hat{\delta})] S_{ev}^{-1} \hat{\xi} \hat{\delta}' S_{a0} - (\hat{\delta}' S_{av} + S_{va} \hat{\delta} + \hat{\delta}' S_{aa} \hat{\delta}) \hat{\delta}'.
$$

$$
= [S_{ea} \hat{\delta} - \hat{\xi} \hat{\delta}' S_{ev} - \hat{\xi} \hat{\delta}' S_{aa} \hat{\delta}] S_{ev}^{-1} \hat{\delta}' S_{ev} - S_{ea} \hat{\delta} \hat{\delta}' - \hat{\delta}' S_{aa} \hat{\delta} \hat{\delta}'.
$$

where we again use (A.2).
Proof of (24). Because \( \hat{\xi} - \hat{\xi} \) and \( \hat{\Omega} - \hat{\Omega} \) are both in \( O\pi(T^{-1}) \) and \( S_{0\pi}^{-1} S_{0\pi} \in O\pi(1) \), we find from (21) and (A.2) that

\[
\delta = S_{a\pi}^{-1}(S_{a\pi} - S_{a\pi} \hat{\xi}^t) \hat{\Omega}^{-1} \hat{\xi} (\hat{\xi}^t \hat{\Omega}^{-1} \hat{\xi})^{-1} + O\pi(T^{-1}S_{a\pi}^{-1/2}) \\
= S_{a\pi}^{-1} S_{\pi x, a} \hat{\Omega}^{-1} \hat{\xi} (\hat{\xi}^t \hat{\Omega}^{-1} \hat{\xi})^{-1} + O\pi(T^{-1}S_{a\pi}^{-1/2}).
\]

This completes the proof of Theorem 2 on the expansion of the estimators.

The expansions can conveniently be expressed in terms of a projection matrix

\[
\hat{P} = P(\hat{\xi}, \hat{\Omega}) = \hat{\xi} (\hat{\xi}^t \hat{\Omega}^{-1} \hat{\xi})^{-1} \hat{\xi}^t \hat{\Omega}^{-1} = P(\xi, \Omega),
\]

because from (24)

\[
\hat{\xi} \delta' = P(\hat{\xi}, \hat{\Omega}) S_{\pi x, a} S_{a\pi}^{-1} + O\pi(T^{-1}S_{a\pi}^{-1/2}). \tag{A.3}
\]

Further we have

\[
S_{\pi x, a} = S_{\pi a} - S_{a\pi} S_{a\pi}^{-1} S_{\pi a} = S_{\pi a} + O\pi(T^{1/2}T^{-1}S_{a\pi}^{-1/2}) = S_{\pi a} + O\pi(T^{-1/2}S_{a\pi}^{-1/2}),
\]

and we find the intermediate results, where \( \kappa_{\xi} = (\hat{\xi}^t \hat{\Omega}^{-1} \hat{\xi})^{-1/2} \)

\[
\hat{\xi} \delta' S_{a\pi} = 0 = \hat{\xi} (\hat{\xi}^t \hat{\Omega}^{-1} \hat{\xi})^{-1} \hat{\xi}^t \hat{\Omega}^{-1} S_{\pi x, a} S_{a\pi}^{-1} S_{a\pi} = P S_{\pi a} S_{a\pi}^{-1} S_{a\pi}
\]

\[
\hat{\xi} \delta' S_{a\pi} \delta' = 0 = P S_{\pi a} S_{a\pi}^{-1} S_{\pi x, a} \hat{\Omega}^{-1} \hat{\xi} (\hat{\xi}^t \hat{\Omega}^{-1} \hat{\xi})^{-1} = P S_{\pi a} S_{a\pi}^{-1} S_{a\pi} \kappa_{\xi}
\]

\[
S_{\pi x, a} \delta' = 0 = S_{\pi a} S_{a\pi}^{-1} S_{\pi x, a} \hat{\Omega}^{-1} \hat{\xi} (\hat{\xi}^t \hat{\Omega}^{-1} \hat{\xi})^{-1} = S_{\pi a} S_{a\pi}^{-1} S_{a\pi} \kappa_{\xi}, \tag{A.4}
\]

\[
\hat{\xi} \delta' S_{\pi x, a} = 0 = P S_{\pi a} S_{a\pi}^{-1} S_{\pi a}
\]

\[
\hat{\xi} \delta' S_{a\pi} \delta' \hat{\xi} = 0 = P S_{\pi a} S_{a\pi}^{-1} S_{a\pi} P'.
\]

We next give an expansion of the projection matrix \( \hat{P} \), which will be used in the detailed calculations that follow. We define

\[
P(\hat{\xi}, \hat{\Omega}) = \hat{\xi} (\hat{\xi}^t \hat{\Omega}^{-1} \hat{\xi})^{-1} \hat{\xi}^t \hat{\Omega}^{-1} = \hat{\Omega}^{-1} - \hat{\Omega}^{-1} \hat{P}(\hat{\xi}, \hat{\Omega}).
\]

Note that we only expand as a function of \( \hat{\xi} \) but keep \( \hat{\Omega} \).

Lemma 11. The expansion of \( P(\hat{\xi}, \hat{\Omega}) \) around the point \( (\xi, \hat{\Omega}) \) is given by

\[
P(\hat{\xi}, \hat{\Omega}) = P(\xi, \Omega) + \hat{\Omega}^{-1} \xi (\xi^t \hat{\Omega}^{-1} \xi)^{-1} \xi^t \hat{\Omega}^{-1} \\
+ \hat{\Omega}^{-1} \xi (\xi^t \hat{\Omega}^{-1} \xi)^{-1} (\hat{\xi} - \xi)^t \hat{P}(\xi, \Omega) \\
+ \hat{\Omega}^{-1} (\xi^t \hat{\Omega}^{-1} \xi)^{-1} \hat{\xi} (\xi^t \hat{\Omega}^{-1} \xi)^{-1} (\xi^t \hat{\Omega}^{-1} \xi)^{-1} \xi^t \hat{P}(\xi, \Omega) \\
- \hat{\Omega}^{-1} \xi (\xi^t \hat{\Omega}^{-1} \xi)^{-1} (\hat{\xi} - \xi)^t \hat{P}(\xi, \Omega) \\
- \hat{\Omega}^{-1} (\xi^t \hat{\Omega}^{-1} \xi)^{-1} (\hat{\xi} - \xi)^t \hat{P}(\xi, \Omega) \\
- \hat{\Omega}^{-1} \xi (\xi^t \hat{\Omega}^{-1} \xi)^{-1} (\hat{\xi} - \xi)^t \hat{P}(\xi, \Omega) \\
- \hat{\Omega}^{-1} (\xi^t \hat{\Omega}^{-1} \xi)^{-1} (\hat{\xi} - \xi)^t \hat{P}(\xi, \Omega).
The quadratic term is 

$$P(\xi, \hat{\Omega}) = \xi (\xi \hat{\Omega}^{-1} \xi)^{-1} \xi' \hat{\Omega}^{-1} = \hat{\Omega}^{1/2} u (u' u)^{-1} u' \hat{\Omega}^{-1/2} = \hat{\Omega}^{1/2} P_0 \hat{\Omega}^{-1/2},$$

say, and

$$\bar{P}(\xi, \hat{\Omega}) = \hat{\Omega}^{-1} - \hat{\Omega}^{-1} P(\xi, \hat{\Omega}) = \hat{\Omega}^{-1} - \hat{\Omega}^{-1/2} u (u' u)^{-1} u' \hat{\Omega}^{-1/2} = \hat{\Omega}^{-1/2} \bar{P}_0 \hat{\Omega}^{-1/2}.$$

Then we find using \(\bar{u} = u (u' u)^{-1},\) such that \(\bar{u} u' = P_0\)

$$\hat{\Omega}^{-1/2} P(\xi, \hat{\Omega}) \hat{\Omega}^{1/2} = (u + v)' (u + v) (u + v)' (u + v) = \bar{u} v' (I_n - \bar{u} u') + (I_n - \bar{u} u') v \bar{u}'.$$

The quadratic term is

$$L_2 = -\bar{u} u' v + v \bar{u}' - \bar{u} (v' u + v ' u) \bar{u}' = \bar{u} v' (I_n - \bar{u} u') + (I_n - \bar{u} u') v \bar{u}'$$

The expansion of \(P(\hat{\xi}, \hat{\Omega})\) around the point \((\xi, \hat{\Omega})\), using \(\hat{\xi} - \xi = S_{uv} S_{uv}^{-1},\) \(\kappa \xi = (\xi' \Omega^{-1} \xi)^{-1/2},\) and \(S_{uv} P(\xi, \Omega) S_{uv} = S_{ub} S_{bu},\) we find the following corollary.

**COROLLARY 12.** The expansion of \(P(\xi, \hat{\Omega})\) around the point \((\xi, \hat{\Omega})\), using \(\hat{\xi} - \xi = S_{uv} S_{uv}^{-1},\) \(\kappa \xi = (\xi' \Omega^{-1} \xi)^{-1/2},\) and \(S_{uv} P(\xi, \Omega) S_{uv} = S_{ub} S_{bu},\) is given by

$$P(\xi, \hat{\Omega}) \frac{1}{\hat{\Omega}} P(\xi, \hat{\Omega}) S_{uv} S_{uv}^{-1} (\xi' \hat{\Omega}^{-1} \xi)^{-1} \xi' \hat{\Omega}^{-1}$$

$$+ \hat{\Omega}^{-1} (\xi' \hat{\Omega}^{-1} \xi)^{-1} S_{uv}^{-1} S_{uv} P(\xi, \hat{\Omega})$$

$$+ \xi_1 (\xi' \Omega^{-1} \xi_1)^{-1/2} S_{ub} S_{bu}^{-1} \kappa \xi S_{uv}^{-1} S_{uv} S_{uv}^{-1} (\xi_1' \Omega \xi_1)^{-1/2} \xi_1'$$

$$- \Omega^{-1} (\xi' \Omega^{-1} \xi)^{-1} S_{uv}^{-1} S_{uv} S_{uv}^{-1} (\xi' \Omega^{-1} \xi)^{-1} \xi' \Omega^{-1}$$

$$- \Omega^{-1} (\xi' \Omega^{-1} \xi)^{-1} S_{uv}^{-1} S_{uv} \kappa \xi S_{uv}^{-1} S_{uv} (\xi_1' \Omega \xi_1)^{-1/2} \xi_1'$$

$$- \xi_1 (\xi' \Omega \xi_1)^{-1/2} S_{ub} S_{bu}^{-1} \kappa \xi S_{uv}^{-1} (\xi' \Omega^{-1} \xi)^{-1} \xi' \Omega^{-1}.$$
Proof of Theorem 3

We start with (25):

\[-2 \log LR = \text{tr} \{Q_0\} + T^{-1}(\text{tr} \{Q_1\} + \frac{1}{2} \text{tr} \{Q_0^2\})\]

and evaluate each term starting with the ones that have a factor \(T^{-1}\).

Calculation of \(\text{tr} \{Q_1\}\). We find from Theorem 2 that

\[\text{tr} \{Q_1\} = T \text{tr} \{\hat{\Omega}^{-1} [S_{e,a} \delta - \hat{\xi} \hat{\delta} S_{a} - \hat{\xi} \hat{\delta} S_{a} \delta] S_{v}^{-1} [\hat{\delta} S_{a,e} - S_{v} \hat{\delta} \hat{\xi} - \hat{\delta} S_{a} \hat{\xi}'] \}\]

Because \(\text{tr} \{Q_1\}\) is multiplied by \(T^{-1}\) we need only retain the main term in each of the matrices. The key quantities are calculated in (A.4), and we find

\[S_{e,a,e} \delta - \hat{\xi} \hat{\delta} S_{a} - \hat{\xi} \hat{\delta} S_{a} \delta = (I_n - P)S_{e,a}^{-1} S_{a} \kappa \hat{\xi} - PS_{_{e,a}} S_{a}^{-1} S_{a},\]

such that with \(\hat{\Omega}\) replaced by \(\Omega\)

\[\text{tr} \{Q_1\} = T \text{tr} \{\kappa \xi S_{a} S_{a}^{-1} S_{a} \Omega^{-1} (I_n - P) S_{e,a}^{-1} S_{a} \xi \Omega S_{v}^{-1}\}
\]

\[\quad + S_{a} S_{a}^{-1} S_{a} \Omega^{-1} P S_{_{e,a}} S_{a}^{-1} S_{a} S_{v}^{-1}\}
\]

\[\quad = T \text{tr} \{\kappa \xi S_{a} S_{a}^{-1} S_{a} \xi \Omega S_{v}^{-1}\}
\]

\[\quad + T \text{tr} \{S_{a} S_{a}^{-1} S_{a} \xi \Omega S_{v}^{-1}\},\]  \(\text{(A.5)}\)

where the expression has been reduced using the properties of the projection matrix

\[P' \Omega^{-1} (I_n - P) = 0, \quad P' \Omega^{-1} P = P' \Omega^{-1}\]

\[(I_n - P)' \Omega^{-1} (I_n - P) = \Omega^{-1} (I_n - P) = \xi \xi^{-1} \xi^{-1} \xi^{-1}\]

and the identities (see (18) for the definition of \(U_i\) and \(B_i\))

\[S_{e} \Omega^{-1} P S_{e} = S_{e} \Omega^{-1} \xi \xi^{-1} \xi^{-1} \xi^{-1} S_{e} = S_{a} S_{a}\]

\[S_{e} \Omega^{-1} (I_n - P) S_{e} = S_{e} \Omega^{-1} \xi \xi^{-1} \xi^{-1} \xi^{-1} S_{e} = S_{e} S_{e}\]  \(\text{(A.6)}\)

Calculation of \(\text{tr} \{Q_0^2\}\). We find from Theorem 2 that from (A.4)

\[Q_0 = \hat{\Omega}^{-1} [S_{e,a,e} \delta + \hat{\xi} \hat{\delta} S_{a,e} - \hat{\xi} \hat{\delta} S_{a} \delta]\]

\[\quad = \Omega^{-1} [(I_n - P) S_{e,a} S_{a}^{-1} S_{a} P' + PS_{_{e,a}} S_{a}^{-1} S_{a}]\]

and hence

\[Q_0^2 = \Omega^{-1} (I_n - P) S_{e,a} S_{a}^{-1} S_{a} P' \Omega^{-1} P S_{_{e,a}} S_{a}^{-1} S_{a}\]

\[\quad + \Omega^{-1} P S_{_{e,a}} S_{a}^{-1} S_{a} P' \Omega^{-1} (I_n - P) S_{e,a} S_{a}^{-1} S_{a} P'\]

\[\quad + \Omega^{-1} P S_{_{e,a}} S_{a}^{-1} S_{a} \Omega^{-1} P S_{_{e,a}} S_{a}^{-1} S_{a},\]
using $P'\Omega^{-1}(I_n - P) = 0$. Hence again using (A.6),
\[ \frac{1}{2}\text{tr}\{Q_0^2\} = \frac{1}{2}\text{tr}\{(S_{aa}^{-1}S_{aa}^{-1})^2\} + \text{tr}\{S_{aa}^{-1}S_{ab}^{-1}S_{ba}^{-1}S_{aa}^{-1}\}. \tag{A.7} \]

The main term $\text{tr}\{Q_0\}$. This term is of the order of one, and hence we have to keep more terms in the expansions. From (A.3) we find that $\hat{\delta}'$ is of the order of $S_{aa}^{-1/2}$ and that
\[ \hat{\delta}' = \hat{P}_S S_{aa}^{-1} + T^{-1}\hat{\delta}'_1 + o_p(T^{-1}S_{aa}^{-1/2}), \]
for some $\hat{\delta}'_1 \in O_p(S_{aa}^{-1/2})$, and hence
\[ \hat{\delta}' S_{aa}^{-1} = \hat{P}_S S_{aa}^{-1} S_{aa}^{-1} S_{aa}^{-1} + T^{-1}\hat{\delta}'_1 S_{aa}^{-1} \]
We first want to show that the coefficient $\hat{\delta}'_1$ does not play a role because of the way the projection matrices enter, such that we can replace $\hat{\delta}'$ by $\hat{P}_S S_{aa}^{-1}$, introducing errors of at most the order $o_p(T^{-1}S_{aa}^{-1/2})$ in the expression for $Q_0$. We find from
\[ \hat{\Omega}Q_0(\hat{\delta}'_1) = S_{aa}^{-1} \hat{\delta}'_1 + \hat{\delta}' S_{aa}^{-1} \hat{\delta}'_1 \]
that
\[ \hat{\Omega}[Q_0(\hat{\delta}'_1) - Q_0(\hat{P}_S S_{aa}^{-1})] = T^{-1}[S_{aa}^{-1} \hat{\delta}'_1 \hat{\delta}'_1 + \hat{\delta}'_1 S_{aa}^{-1} \hat{\delta}'_1 - \hat{\delta}'_1 S_{aa}^{-1} S_{aa}^{-1} \hat{\delta}'_1 - \hat{\delta}'_1 S_{aa}^{-1} \hat{\delta}'_1] \]
\[ = T^{-1}[(I_n - \hat{P}) S_{aa}^{-1} \hat{\delta}'_1 \hat{\delta}'_1 + \hat{\delta}'_1 S_{aa}^{-1} (I_n - \hat{P})^T]. \]
This term, however, does not give a contribution because
\[ \text{tr}\{Q_0\} = \frac{1}{2}\text{tr}\{\hat{\Omega}^{-1}(I_n - \hat{P}) S_{aa}^{-1} \hat{\delta}'_1 \hat{\delta}'_1\} = \text{tr}\{\hat{\Omega}^{-1}(I_n - \hat{P}) S_{aa}^{-1} \hat{\delta}'_1\} \]
and $\hat{\Omega}^{-1}(I_n - \hat{P}) = 0$. In the following we therefore replace $\hat{\delta}'_1$ by $S_{aa}^{-1} S_{aa}^{-1} \hat{\delta}'_1$, and we find
\[ \text{tr}\{Q_0\} = \frac{1}{2}\text{tr}\{\hat{\Omega}^{-1} S_{aa}^{-1} S_{aa}^{-1} \hat{\delta}'_1 \hat{\delta}'_1 + \text{tr}\{\hat{\Omega}^{-1} \hat{P}_S S_{aa}^{-1} S_{aa}^{-1} (I - \hat{P})^T\} \]
where we have used that $\hat{P}_S \hat{\Omega}^{-1}(I_n - \hat{P}) = 0$.

We next expand around $\hat{\xi}$ but keep $\hat{\Omega}$. We find from Corollary 12, using the notation $\hat{\Omega}(\xi, \hat{\Omega}) = \xi_0 (\xi_0^T \hat{\Omega} \xi_0)^{-1} \xi_0$, that
\[ \text{tr}\{Q_0\} = \frac{1}{2}\text{tr}\{S_{aa}^{-1} S_{aa}^{-1} \hat{\Omega}^{-1} P(\hat{\xi}, \hat{\Omega}) S_{aa}^{-1} \}
\[ + 2 \text{tr}\{S_{aa}^{-1} S_{aa}^{-1} P(\hat{\xi}, \hat{\Omega}) S_{aa}^{-1} S_{aa}^{-1} (\xi' \hat{\Omega}^{-1} \xi)' - \xi' \hat{\Omega}^{-1} \xi\}
\[ + \text{tr}\{S_{aa}^{-1} S_{aa}^{-1} S_{aa}^{-1} S_{aa}^{-1} S_{aa}^{-1} (\xi_0^T \hat{\Omega} \xi_0)^{-1} \xi_0^T \hat{\Omega}^{-1} \xi_0\}
\[ - \text{tr}\{S_{aa}^{-1} S_{aa}^{-1} S_{aa}^{-1} S_{aa}^{-1} S_{aa}^{-1} S_{aa}^{-1} (\xi_0^T \hat{\Omega} \xi_0)^{-1} \xi_0^T \hat{\Omega}^{-1} \xi_0\}
\[ - 2 \text{tr}\{S_{aa}^{-1} S_{aa}^{-1} S_{aa}^{-1} S_{aa}^{-1} S_{aa}^{-1} S_{aa}^{-1} (\xi_0^T \hat{\Omega} \xi_0)^{-1} \xi_0^T \hat{\Omega}^{-1} \xi_0\} = A_1 + A_2 + A_3 + A_4 + A_5. \]
The last three terms are of the order $T^{-1}$ and simplify using the relations
\begin{align*}
S_{a,u,v} \Omega_{-1} \xi (\xi' \Omega_{-1} \xi')^{-1} & = S_{a,u} + O_p(S_{aa}^{1/2} T^{-1/2}), \\
S_{a,u,v} \xi_{\perp} (\xi'_{\perp} \Omega_{\perp} \xi_{\perp})^{-1/2} & = S_{ab,u} + O_p(S_{aa}^{1/2} T^{-1/2}).
\end{align*}

We find
\begin{align*}
A_3 & \equiv \frac{1}{2} \text{tr} \{ S_{a,u}^{-1} S_{ab} S_{vu}^{-1} \kappa_{\xi}^2 S_{uv}^{-1} S_{ub} S_{ba} \}, \\
A_4 & \equiv - \frac{1}{2} \text{tr} \{ S_{a,u}^{-1} S_{ab} \kappa_{\xi} S_{uv}^{-1} S_{vu} S_{uv}^{-1} \kappa_{\xi} S_{ua} \}, \\
A_5 & \equiv - 2 \text{tr} \{ S_{a,u}^{-1} S_{ab} S_{vu}^{-1} \kappa_{\xi} S_{vu} S_{uv}^{-1} \kappa_{\xi} S_{ua} \}.
\end{align*}
(A.8)

We still need to simplify the first two terms, $A_1$ and $A_2$. We introduce the definitions
\begin{align*}
L' & = \begin{pmatrix}
(\xi' \Omega_{-1} \xi)^{-1/2} \xi' \\
(\xi_{\perp} \Omega_{\perp} \xi_{\perp})^{-1/2} \xi_{\perp}'
\end{pmatrix}, & L'e_i & = \begin{pmatrix}
U_i \\
B_i
\end{pmatrix},
\end{align*}
and from $\hat{\Omega} = T^{-1} S_{\epsilon \epsilon, \nu}$, we find
\begin{align*}
L' \hat{\Omega} L & = T^{-1} \begin{pmatrix}
S_{a,u,v} & S_{ab,u} \\
S_{ab,u} & S_{bb,v}
\end{pmatrix}, & S_{a,u,v} L & = \begin{pmatrix}
S_{a,u,v} \\
S_{ab,u}
\end{pmatrix}
\end{align*}
such that
\begin{align*}
S_{a,u,v} \hat{\Omega}^{-1} \xi & = S_{a,u,v} L (L' \hat{\Omega} L)^{-1} L' \xi \\
& = T (S_{a,u,v}^{-1} S_{ab,u}) \begin{pmatrix}
S_{a,u,v} & S_{ab,u} \\
S_{ab,u} & S_{bb,v}
\end{pmatrix}^{-1} \begin{pmatrix}
(\xi' \Omega_{-1} \xi)^{1/2} \\
0
\end{pmatrix} \\
& = TS_{a,u,v,b}^{-1} (\xi' \Omega_{-1} \xi)^{1/2},
\end{align*}
and
\begin{align*}
(\xi' \Omega_{-1} \xi)^{1/2} (\xi' \hat{\Omega}^{-1} \xi)^{-1} (\xi' \Omega_{-1} \xi)^{1/2} & = [(\xi' \Omega_{-1} \xi)^{-1/2} \xi' L (L' \hat{\Omega} L)^{-1} L' \xi (\xi' \Omega_{-1} \xi)^{-1/2}]^{-1} \\
& = T^{-1} \begin{pmatrix}
(I_n) & S_{a,u,v} \\
0 & S_{ab,u}
\end{pmatrix}^{-1} \begin{pmatrix}
(I_n) \\
0
\end{pmatrix}^{-1} = T^{-1} S_{a,u,v,b},
\end{align*}
which combine to give
\begin{align*}
S_{a,u,v} \hat{\Omega}^{-1} \xi (\xi' \hat{\Omega}^{-1} \xi)^{-1} = S_{a,u,v,b} \kappa_{\xi}
\end{align*}
(A.9)
and hence
\[ A_1 = \text{tr} \{ S_{aa}^{-1} S_{ae,v} \hat{\Omega}^{-1} P(\xi, \hat{\Omega}) S_{e.a.v} \} \]
\[ = \text{tr} \{ S_{aa}^{-1} [ S_{ae,v} \hat{\Omega}^{-1} \xi] [(\xi' \hat{\Omega}^{-1} \xi)^{-1}] \xi' \hat{\Omega}^{-1} S_{e.a.v} \} \]
\[ = T \text{tr} \{ S_{aa}^{-1} S_{au,b,v} S_{aa}^{-1} S_{au,v,b} \}. \]

Now we want to replace \( S_{aa}^{-1} \) by \( S_{aa,b,v}^{-1} \) and use the identity
\[ S_{aa,b,v} = S_{aa} - S_{ab} S_{bb}^{-1} S_{ba} - S_{aa} - S_{ab} S_{bb}^{-1} S_{ba} - S_{av,b} S_{av,b} \]
to find
\[ S_{aa}^{-1} - S_{aa,b,v}^{-1} = S_{aa}^{-1} - (S_{aa} - S_{ab} S_{bb}^{-1} S_{ba} - S_{av,b} S_{av,b})^{-1} \]
\[ = -S_{aa}^{-1} (S_{ab} S_{bb}^{-1} S_{ba} + S_{av,b} S_{av,b}) S_{aa}^{-1} + O_p(S_{aa}^{-1} T^{-2}), \]
such that for \( A_1 \) we find the main term \( T \text{tr} \{ S_{aa,b,v}^{-1} S_{au,b,v} S_{aa,b,v}^{-1} S_{au,v,b} \} \).

The last two terms are of the order of \( T^{-1} \) and can be simplified using
\[ T^{-1} S_{bb} \sim I_n, \quad S_{av,b} \sim S_v, \quad S_{au,b,v} \sim S_{aa}, \quad T^{-1} S_{uu,b,v} \sim I_n, \]
\[ T^{-1} S_{uv,b} \sim T^{-1} S_{uv}, \]
and we find
\[ A_1 = T \text{tr} \{ S_{aa,b,v}^{-1} S_{au,b,v} S_{aa,b,v}^{-1} S_{au,v,b} \} \]
\[ - T^{-1} \text{tr} \{ S_{aa}^{-1} S_{ab} S_{bb}^{-1} S_{ba} S_{aa} \} - \text{tr} \{ S_{aa}^{-1} S_{au} S_{uu}^{-1} S_{uv} S_{au}^{-1} S_{aa} \}. \]  \hspace{1cm} (A.10)

To find \( A_2 \) we note that \( S_{ae,v} \xi \perp = S_{ab,v} (\xi_1' \Omega \xi_1)^{1/2} \) and that
\[ (\xi_1' \Omega \xi_1)^{1/2} (\xi_1' \hat{\Omega} \xi_1)^{-1} (\xi_1' \Omega \xi_1)^{1/2} = \left[ (\xi_1' \Omega \xi_1)^{-1/2} \xi_1 S_{ee,v} \xi_1 (\xi_1' \Omega \xi_1)^{-1/2} \right]^{-1} = T S_{bb,v}, \]
which combine to give
\[ S_{ae,v} \xi \perp (\xi_1' \hat{\Omega} \xi_1)^{-1/2} S_{uv} = T S_{ab,v} S_{bb,v} S_{bu}. \]  \hspace{1cm} (A.11)

We then find
\[ A_2 = T \text{tr} \{ S_{aa}^{-1} S_{ae,v} \xi \perp (\xi_1' \hat{\Omega} \xi_1)^{-1} S_{uv} S_{ev}^{-1} (\xi_1' \hat{\Omega}^{-1} \xi_1)^{-1} \xi_1' \hat{\Omega}^{-1} S_{e.a.v} \} \]
\[ = 2 T \text{tr} \{ S_{aa}^{-1} S_{ab,v} S_{bb,v} S_{uv}^{-1} S_{bu} \xi \perp S_{au,v,b} \}, \]
using (A.9) and (A.11). Again we can use the relation \( T^{-1} S_{bb,v} = T^{-1} S_{bb} + O_p(T^{-1}) \) to simplify so that we get
\[ A_2 = 2 T \text{tr} \{ S_{aa}^{-1} S_{ab,v} S_{bb} S_{bu} S_{uv}^{-1} \xi \perp S_{au,v,b} \}. \]  \hspace{1cm} (A.12)

Inserting the expressions (A.5), (A.7), (A.8), (A.10), and (A.11) into the expression for the likelihood ratio test (25) we find the expression in Theorem 3. Note that two
terms from $A_1$, as given in (A.10), cancel a term in the expression for $Q_1$ (A.5) and one in the expression for $Q_0$ (A.7).

This completes the proof of the representation of the likelihood ratio test statistic given in Theorem 3.

**Proof of Theorem 4**

Throughout we need the following result from the multivariate normal distribution.

**LEMMA 13.** Let $X$ and $Y$ be multivariate normally distributed with mean zero and covariance matrix $\Sigma \otimes \Omega$, that is,

$$E[\lambda' X \lambda' Y \lambda_3] = \lambda_1' \Sigma \lambda' \lambda_3 \lambda_2' \Omega \lambda_4.$$  

Then, for $A$ and $B$ of suitable dimensions, we have

$$E[XY'] = \Sigma \text{tr} \{A'\}, \quad E[XY] = \Sigma' \Omega.$$  

**Proof.** We let $A = \sum_{i=1}^{\infty} \mu_i \nu_i'$, such that

$$E[\mu' X Y' \nu] = \sum_{i=1}^{\infty} E[\mu' X \mu_i \nu' Y \nu_i] = \mu' \Sigma \nu \sum_{i=1}^{\infty} \mu_i \nu_i' = \mu' \Sigma \nu \text{tr} \{A'\}.$$  

Next let $B = \sum_{i=1}^{\infty} \eta_i \xi_i'$; then

$$E[\mu' X Y \nu] = \sum_{i=1}^{\infty} E[\mu' X \eta_i \xi_i' Y \nu] = \sum_{i=1}^{\infty} \mu' \Sigma \eta_i \eta_i' \nu = \mu' \Sigma' B' \Omega \nu.$$  

We also need the calculation given in the following lemma.

**LEMMA 14.** Let $S_t = \sum_{i=0}^{\infty} \sigma_i U_{t-i}$ and $H_t = \sum_{i=0}^{\infty} \eta_i U_{t-i}$, where $\sigma_i$ and $\eta_i$ are coefficients matrices that decrease exponentially fast. Let $f_t$ and $g_t$ be sequences of vectors; then

$$E[\lambda' M_f \lambda' M_h \lambda_3] = \lambda_1' E\left[ \sum_{i,j} f_i U_{t-i} \sigma_j' \lambda_2' \lambda_3' g_j U_{t-j} \eta_j' \lambda_4 \right]$$

$$= \sum_{i,j} \lambda_1' f_i g_{t-i-j} \lambda_2' \sigma_j' \lambda_3' \eta_j' \lambda_4.$$  

(A.13)

For suitable choices of coefficients we can obtain results about product moments such as $M_{uu}, M_{by}$, etc. Subsequently we indicate by $E_{\xi_i} [...]$ the conditional expectation and leave out the conditioning variables $\xi_i' \varepsilon_i$. Notice that when we condition on $\xi_i' \varepsilon_i$, the processes $A_{t-1}$ and $B_t$ are fixed and denoted by $a_{t-1}$ and $b_t$. This also holds in the case of (14), where we condition on $\alpha_{t-1} \varepsilon_t$, rather than $\alpha_t \varepsilon_t$.

To systematize the calculations in the proof of Theorem 4 we define some linear functions

$$N_{ij} = M_{ij}^{1/2} (M_{ij} - E_{\xi_i} (M_{ij})), \quad i = a, b, d, \quad j = u, y, \quad \mu_{bT} = E_{\xi_n} (M_{by})$$

and some quadratic functions

$$Q_{yy} = T^{-1/2} (M_{yy} - T \Sigma), \quad Q_{yu} = T^{-1/2} M_{yu}.$$
LEMMA 15. Under Assumption 1, and conditional on $\xi^1$, e, it holds that all the linear and quadratic functions, $N_\theta$, and $Q_\theta$, are asymptotically normally distributed with mean zero. The linear functions are asymptotically independent (a.i.) of the quadratic functions. Moreover we find that $N_{by}$ is a.i. of $(N_{au}, N_{bu}, N_{ay}, N_{dy})$ and $N_{ay}$ is a.i. of $N_{yd}$. Subsequently we give some results that are needed in the proof of Theorem 4.

\[
\text{Var}_\xi(N_{au}) = I_{n_a} \otimes I_{n_a},
\]
\[
\text{Cov}_\xi(N_{au}, N_{ay}) = I_{n_a} \otimes \theta',
\]
\[
E_\xi[N_{ay} N_{au}] = n_a \theta,
\]
\[
\text{Var}_\xi(N_{by}) = I_{n_b} \otimes \Sigma_{\theta \theta},
\]
\[
T^{-1} \mu_{by} A \mu_{by} = \text{tr}(A) \Sigma_{\phi \phi},
\]
\[
E_\xi[M_{y\theta}] = \Sigma,
\]
\[
T^{-1} E_\xi[M_{y\theta} M_{by}] = n_b \Sigma,
\]
\[
T^{-1} E_\xi[M_{y\theta} \Sigma^{-1} \kappa^\prime \Sigma^{-1} M_{y\theta}] = I_{n_b} \text{tr}\{\kappa^\prime \Sigma^{-1} \kappa\}
\]

Proof. It follows from the central limit theorem that all variables are asymptotically normal. We shall prove the remaining statements by calculating variances and covariances, using Lemmas 13 and 14.

We find immediately the result (A.14), because $M_{au}$ is Gaussian with mean zero and variance $M_{au} \otimes I_{n_a}$. To prove (A.15) we find from Lemma 14 with $f_t = g_t = a_{r-1}, S_t = U_t, H_t = Y_{r-1}$

\[
\text{Cov}(\lambda' N_{au} \lambda_2, \lambda' N_{ay} \lambda_4) = \sum_t \lambda_1' M_{uu}^{-1/2} \sum_t a_{r-1} a_{r-i} M_{uu}^{-1/2} \lambda_3 \lambda_2' \theta' \lambda_4
\]
\[
= \lambda_1' \lambda_3 \lambda_2' \theta' \lambda_4,
\]

where we have used (26). Relation (A.16) follows from (A.15) and Lemma 13.

To prove (A.17) we use Lemma 14 with $f_t = g_t = b_t$ and $S_t = H_t = Y_{r-1}$:

\[
\text{Var}_\xi(\lambda' N_{by} \lambda_2) = \sum_{i,j} \lambda_1' M_{bb}^{-1/2} \sum_t b_t b_{t+i} M_{bb}^{-1/2} \lambda_1 \text{tr}\{\lambda_2' \theta' \lambda_2\}
\]
\[
= \lambda_1' I_{n_b} \lambda_2' \Sigma_{\theta \theta} \lambda_2.
\]

Here we used (27) and $\Sigma_{\theta \theta} = \sum \theta_i \theta_i'$ (see (31)). The statement (A.18) follows from (28) because

\[
T^{-1} \mu_{by} A \mu_{by} = T^{-1} \sum_{t,s,i} \psi_t b_{t+j} b_{t+i} A b_t' b_{t+i-j} \psi_t' = \sum_i \text{tr}(A) \psi_i \psi_i' = \text{tr}(A) \Sigma_{\phi \phi}.
\]
To prove (A.19) we use Lemma 13, (27), and (31) and find

\[
T^{-1} E_{\xi_1} [M_{yy}] = T^{-1} \sum_{t,i,j} \theta_i E_{\xi_1} [U_{t-i-1} U'_{t-j-1}] \psi_j' + \sum_{t,i,j} \psi_i b_{t-i-1} b'_{t-j-1} \psi_j' \\
= \sum_i \theta_i \psi_i' + \sum_i \psi_i \psi_i' = \Sigma.
\]

Relation (A.20) follows from Lemma 13, (A.17) and (A.18) and

\[
T^{-1} E_{\xi_1} [M_{yb} M_{by}] = T^{-1} E_{\xi_1} [N_{yb} M_{bb} N_{by}] + T^{-1} \mu_{bT} \mu_{bT} \\
= T^{-1} \Sigma_{\theta \theta} \text{tr} \{M_{bb}\} + \text{tr} \{I_{n_b}\} \Sigma_{\phi \phi} = n_b \Sigma.
\]

The result (A.21) follows from (A.17), (A.18), and Lemma 13.

To complete the proof we need to show that some covariances are zero. The basic calculation is given in Lemma 14. We find that

\[
\begin{align*}
\text{Cov}_{\xi_1} [\lambda'_1 M_{dd}^{-1/2} M_{du} \lambda_2, \lambda'_3 M_{bb}^{-1/2} M_{by} \lambda_4] & = \sum_{t,i} \lambda'_1 M_{dd}^{-1/2} b_t b'_{t+i+1} M_{bb}^{-1/2} \lambda_3 \lambda'_2 \theta_i \lambda_4, \\
\text{Cov}_{\xi_1} [\lambda'_1 M_{ma}^{-1/2} M_{mu} \lambda_2, \lambda'_3 M_{bb}^{-1/2} M_{by} \lambda_4] & = \sum_{t,i} \lambda'_1 M_{ma}^{-1/2} a_{t-i} b'_{t-i+j} M_{bb}^{-1/2} \lambda_3 \lambda'_2 \theta_i \lambda_4, \\
\text{Cov}_{\xi_1} [\lambda'_1 M_{bb}^{-1/2} M_{bu} \lambda_2, \lambda'_3 M_{bb}^{-1/2} M_{by} \lambda_4] & = \sum_{t,i} \lambda'_1 M_{bb}^{-1/2} b_t b'_{t+i+1} M_{bb}^{-1/2} \lambda_3 \lambda'_2 \theta_i \lambda_4, \\
\text{Cov}_{\xi_1} [\lambda'_1 M_{ma}^{-1/2} M_{du} \lambda_2, \lambda'_3 M_{dd}^{-1/2} M_{by} \lambda_4] & = \sum_{t,i} \lambda'_1 M_{ma}^{-1/2} a_{t-i} b'_{t-i+j} M_{dd}^{-1/2} \lambda_3 \lambda'_4 \theta_i \lambda_2
\end{align*}
\]

are all \(O(T^{-1/2})\) because of the factor \(M_{bb}^{-1/2}\), and

\[
\text{Cov}_{\xi_1} [\lambda'_1 M_{ma}^{-1/2} M_{mu} \lambda_2, \lambda'_3 M_{dd}^{-1/2} M_{du} \lambda_4] = \sum_{t,i} \lambda'_1 M_{ma}^{-1/2} a_{t-i} d'_{t-i+j} M_{dd}^{-1/2} \lambda_3 \lambda'_4 \theta_i \lambda_2 = 0,
\]

because

\[
\sum_i a_{t-1} d'_{t-i+j} = \sum_i a_{t-1} d_t M^{-i+j} = 0,
\]

as \(a_{t-1}\) has been orthogonalized on \(d_t\). Finally we need the result that the linear variables are asymptotically uncorrelated with the quadratic variables. This follows from the fact that the Gaussian distribution has third moment zero. \(\blacksquare\)
Proof of Theorem 4. We use the result of Theorem 3 in the form

\[ E[-2 \log LR(\xi_{\text{red}})] = 1 \cdot K_1 + T^{-1}(K_2 + K_3 + K_4 + K_5 + K_6 + K_7) \]

and evaluate each in turn.

The Main Terms \( K_1 + T^{-1}K_2 \). We have

\[ K_1 + T^{-1}K_2 = TE_{\xi_i}[\{S_{\text{aux,b}}S_{\text{aux,b}}^{-1}S_{\text{aux,b}}^{-1}S_{\text{aux,b}}^{-1}S_{\text{aux,b}}^{-1}\} + \frac{1}{2} T^{-1}\{S_{\text{aux,b}}S_{\text{aux,b}}^{-1}S_{\text{aux,b}}^{-1}\}] \]

This is the correction term given in Johansen (1999) based upon the regression equation

\[ \tilde{\xi}' \Delta X_t = \omega \xi_{\text{reg}}') \Delta X_t + (\phi_i'(\tilde{\xi}' - \omega \xi_{\text{reg}}') \Psi) (V_{t-1}, Z_{t-1} \Psi) + \delta' A_{t-1} + \tilde{\xi}' \Phi \ d_t + \tilde{\varepsilon}_t \]

for the test that \( \delta = 0 \), when \( \xi \) is known (\( \tilde{\xi} = \xi(\xi' \xi)^{-1} \)). Note that just as \( \xi_{\text{reg}}' \sum_{i=1}^{T-1} e_i \)

\[ \text{enters as a regressor in } A_{t-1}, \text{ one can check that } \xi_{\text{reg}}' \Delta X_t \text{ can be replaced by } B_t = (\xi_{\text{reg}}' \Omega \xi_{\text{reg}})^{-1/2} \xi_{\text{reg}}' e_t, \text{ because (17) shows that } \xi_{\text{reg}}' \Delta X_t \text{ is a linear function of } B_t, \]

and \( (Z_{t-1}, d_t) \), which are corrected for in the regression. The result given there states that

\[
K_1 + T^{-1}K_2 = n_v n_a + n_n n_n + \frac{n_v n_a}{T} \left[ \frac{1}{2} (n_v + n_a + 1) + (n_a + n_v + n) \right]
\]

\[ + \frac{n_v}{T} \left[ (n_a - 1) v(\xi) + 2(c(\xi) + c_d(\xi)) \right], \]

with the coefficients \( v(\xi), c(\xi), \) and \( c_d(\xi) \) given in Theorem 4.

The rest of the proof of Theorem 4 deals with the problem of showing that

\[ K_3 + K_4 + K_5 + K_6 + K_7 = 0 \]

We first consider the terms \( K_4, K_5, K_6, \) and \( K_7 \). Because they have a factor of \( T^{-1} \), we only need to consider the main terms.

The Term \( K_4 = TE_{\xi_i}[\{S_{\text{aux,b}}S_{\text{aux,b}}^{-1}S_{\text{aux,b}}S_{\text{aux,b}}^{-1}S_{\text{aux,b}}^{-1}S_{\text{aux,b}}^{-1}S_{\text{aux,b}}^{-1}S_{\text{aux,b}}^{-1}\}] \). We reintroduce the product moments \( M_{ab} \) and find \( T^{-1}S_{uv} = T^{-1}M_{uv,z}, \) which converges toward \( \text{Var}(V_{t-1}|Z_{t-1}) = M_{uv,z} \), say. We find \( S_{ab} = M_{ab,z},d M_{ab,z},d M_{ab,z},d \) where the highest order term is \( M_{ab}M_{ab}^{-1}M_{ab} \), and we find from (A.14) and Lemma 13 that

\[
K_4 = E_{\xi_i}[\{S_{\text{aux,b}}S_{\text{aux,b}}^{-1}S_{\text{aux,b}}S_{\text{aux,b}}^{-1}S_{\text{aux,b}}^{-1}S_{\text{aux,b}}^{-1}S_{\text{aux,b}}^{-1}S_{\text{aux,b}}^{-1}\}]
\]

\[ = E_{\xi_i}[\{S_{\text{aux,b}}S_{\text{aux,b}}^{-1}S_{\text{aux,b}}^{-1}M_{ab}M_{ab}M_{ab}^{-1}M_{ab}^{-1}\}]
\]

\[ = \{S_{\text{aux,b}}S_{\text{aux,b}}^{-1}S_{\text{aux,b}}^{-1}\} \text{tr}(M_{ab}M_{ab}M_{ab}^{-1}). \]  \hfill (A.23)

The Term \( K_5 = TE_{\xi_i}[\{S_{\text{aux,b}}S_{\text{aux,b}}^{-1}S_{\text{aux,b}}S_{\text{aux,b}}^{-1}S_{\text{aux,b}}^{-1}S_{\text{aux,b}}^{-1}S_{\text{aux,b}}^{-1}S_{\text{aux,b}}^{-1}\}] \). We replace \( S_{ab}^{-1}S_{ab}^{-1}S_{ab}^{-1}S_{ab}^{-1}S_{ab}^{-1}S_{ab}^{-1}S_{ab}^{-1}S_{ab}^{-1} \) by \( M_{ab,z},d M_{ab,z},d M_{ab,z},d \). We next consider \( S_{\text{aux,b}}S_{\text{aux,b}}^{-1}S_{\text{aux,b}}^{-1}S_{\text{aux,b}}^{-1}S_{\text{aux,b}}^{-1}S_{\text{aux,b}}^{-1}S_{\text{aux,b}}^{-1}S_{\text{aux,b}}^{-1}\). Because \( Y_i = (V_i', Z_i') \), we have the identity
that we can replace

\[ \kappa_{\xi} S_{uv}^{-1} S_{vb} = \kappa_{\xi} (I_{n_t}, 0_{n_t, n_t}) M_{y, d}^{-1} M_{vb, d} = \tilde{\kappa}_{\xi}^{2} M_{y, d}^{-1} M_{vb, d} \]

(see (34)). We find that \( T^{-1} M_{y, d} \) converges toward \( \text{Var}(Y) = \Sigma \) (see (A.19)), such that we can replace \( \kappa_{\xi} S_{uv}^{-1} S_{vb} \) by \( T^{-1} \tilde{\kappa}_{\xi}^{2} \Sigma^{-1} M_{vb, d} \) because \( M_{vb, d} = M_{vb} + O_{p}(1) \). We find from (A.21) that

\[ K_{S} = T^{-1} E_{\xi_{\ell}} \left[ \text{tr} \{ M_{bb} M_{aa}^{-1} M_{ab} M_{by} \Sigma^{-1} \tilde{\kappa}_{\xi}^{2} \Sigma^{-1} M_{vb, d} \} \right] \]

\[ = 0 \]

**Thus**

\[ K_{S} = T^{-1} E_{\xi_{\ell}} \left[ \text{tr} \{ M_{bb} M_{aa}^{-1} M_{ab} M_{by} \Sigma^{-1} \tilde{\kappa}_{\xi}^{2} \Sigma^{-1} M_{vb, d} \} \right] = 0 \]

\[ K_{S} = T^{-1} E_{\xi_{\ell}} \left[ \text{tr} \{ M_{bb} M_{aa}^{-1} M_{ab} M_{by} \Sigma^{-1} \tilde{\kappa}_{\xi}^{2} \Sigma^{-1} M_{vb, d} \} \right] = 0 \]

**The Term** \( K_{6} = -T E_{\xi_{\ell}} [\text{tr} \{ S_{aa} S_{aa}^{-1} S_{aa} \kappa_{\xi} S_{uu}^{-1} S_{bb} S_{aa}^{-1} \kappa_{\xi} \}]. \) We replace \( S_{aa} = M_{aa, z, d} \) by \( M_{aa} \), and from (A.24) we have that \( S_{uu}^{-1} S_{bb} \) can be replaced by \( T^{-1} (I_{n_s}, 0_{n_s, n_s}) \Sigma^{-1} M_{vb, d} \). We find because \( E_{\xi_{\ell}} [N_{aa} N_{aa}] = n_{a} I_{n_{a}} \) (see (A.14) and Lemma 13) and \( T^{-1} E_{\xi_{\ell}} [M_{vb, d}] = n_{b} \Sigma \) (see (A.20)), that because \( N_{aa} \) and \( N_{by} \) are asymptotically independent we get

\[ K_{6} = 0 \]

**The Term** \( K_{7} = -2T E_{\xi_{\ell}} [\text{tr} \{ S_{aa} S_{aa}^{-1} S_{aa} \kappa_{\xi} S_{uu}^{-1} S_{bb} \}]. \) We find as before that we can use (A.24) to replace \( \kappa_{\xi} S_{uu}^{-1} S_{bb} \) by \( T^{-1} \tilde{\kappa}_{\xi}^{2} \Sigma^{-1} M_{vb, d} \) and by a similar argument \( \kappa_{\xi} S_{uu}^{-1} S_{bb} \) by \( T^{-1} \tilde{\kappa}_{\xi}^{2} \Sigma^{-1} M_{vb, d} \) and find

\[ K_{7} = 0 \]

as \( (N_{aa}, N_{ya}) \) and \( N_{by} \) are independent asymptotically.

**The Term** \( K_{8} = 2T^{2} E_{\xi_{\ell}} [\text{tr} \{ S_{ba, u} S_{aa}^{-1} S_{aa, b, u} \kappa_{\xi} S_{uu}^{-1} S_{vb} \}]. \) The stochastic variable is of the order \( T^{1/2} \), and we need some more care, but it turns out to have an expectation that is of the order of a constant. We can replace a matrix by another, only if the relative error is \( O_{p}(T^{-1}) \). We want to replace \( S_{aa} \) by \( M_{aa} \) and expand as follows:

\[ S_{aa} = M_{aa, z, d} = M_{aa, d} - M_{az, d} M_{zz, d}^{-1} M_{za, d} = M_{aa} - M_{ac} M_{zz, d}^{-1} M_{ca}, \]

because \( M_{ad} = 0 \). Thus \( M_{aa}^{-1} (S_{aa} - M_{aa}) \in O_{p}(T^{-1}) \). Similarly we find

\[ S_{bb} = M_{bb, z, d} = M_{bb, d} - M_{bz, d} M_{zz, d}^{-1} M_{zb, d} = M_{bb} - O_{p}(1) \]
and $M_{bb}^{-1} (S_{bb} - M_{bb}) \in Op(T^{-1})$, such that we can replace $S_{bb}$ and $S_{aa}$ by $M_{bb}$ and $M_{aa}$.

Next we expand using $M_{ad} = 0$:

$$S_{au,v} = M_{au} - M_{ab} M_{bb}^{-1} M_{ba,d} - M_{ay,b} M_{yy,b} M_{yy,b,d}, \quad (A.28)$$

$$S_{ba,v} = M_{ba,y,d} = M_{ba} - M_{by,d} M_{yy,d} M_{ya}, \quad (A.29)$$

$$T^{-1} M_{yy,d} = (\Sigma - (\Sigma - T^{-1} M_{yy,d}^{-1}))^{-1} \frac{1}{2} \Sigma^{-1} + \Sigma^{-1} (\Sigma - T^{-1} M_{yy}) \Sigma^{-1}. \quad (A.30)$$

We insert from (A.24) $\kappa_{\xi} S_{uv}^{-1} S_{ib} = \tilde{\kappa}_{\xi} M_{yy,d}^{-1} M_{ib} d$ and use (A.28–A.30) to find

$$2T^2 S_{au,v} S_{aa}^{-1} S_{au,v,b} \kappa_{\xi} S_{uv}^{-1} S_{ib} S_{bb}^{-1} \kappa_{\xi}^2 M_{yy,d}^{-1} M_{ib}$$

$$= 2T^2 M_{ba,y,d} M_{aa}^{-1} M_{au,y,b,d} \tilde{\kappa}_{\xi} M_{yy,d}^{-1} M_{yb,d} M_{bb}^{-1}$$

$$- 2T^2 M_{by} M_{aa}^{-1} M_{au} \tilde{\kappa}_{\xi} M_{yy,d}^{-1} M_{yb} - 2T^2 M_{ba} M_{aa}^{-1} M_{ab} M_{ba,d} \tilde{\kappa}_{\xi} M_{yy,d}^{-1} M_{yb}$$

$$+ 2M_{ba} M_{aa}^{-1} M_{ab} \tilde{\kappa}_{\xi} M_{yy,d}^{-1} M_{ib}$$

where $T^{-1} M_{bb}$ has been replaced by $I_{aa}$ in the last four terms, because they are of the order of a constant. In this way $K_3$ is split into five terms:

$$K_3 = K_{31} + K_{32} + K_{33} + K_{34} + K_{35}.$$ 

This gives a number of contributions, which we investigate one by one. The term $K_{31}$ is found from

$$K_{31} = 2TE_{\xi} \left[ \{M_{ba} M_{aa}^{-1} M_{aa} \tilde{\kappa}_{\xi} M_{yy,d}^{-1} M_{bb}^{-1} \} \right]$$

$$= 2T E_{\xi} \left[ \{M_{ba} M_{aa}^{-1} M_{aa} \tilde{\kappa}_{\xi} M_{yy,d}^{-1} M_{bb}^{-1} \} \right]$$

$$- 2T E_{\xi} \left[ \{M_{ba} M_{aa}^{-1} M_{aa} \tilde{\kappa}_{\xi} M_{yy,d}^{-1} M_{bb}^{-1} \} \right].$$

The first term is of the order $T^{1/2}$, so asymptotic considerations are not enough and we calculate the mean directly. We find

$$E_{\xi} \left[ M_{aa}^{-1/2} M_{aa} \tilde{\kappa}_{\xi} M_{yy,ib} \right] = E_{\xi} \left[ M_{aa}^{-1/2} \sum_{t,i} a_{t-1} U_t \tilde{\kappa}_{\xi} M_{yy,ib} \right]$$

$$= \sum_{t,i} M_{aa}^{-1/2} a_{t-1} b_{i,i+1} \text{tr} \{ \tilde{\kappa}_{\xi} \Sigma^{-1} \theta \}$$

$$= -M_{aa}^{-1/2} M_{ab} \text{tr} \{ \tilde{\kappa}_{\xi} \Sigma^{-1} \tilde{\kappa}_{\xi} \},$$

because $\theta = -\tilde{\kappa}_{\xi}$ (see (32) and (34)). The second term is zero, because it is of the order of a constant, and $N_{aa}$ and $N_{yd}$ are asymptotically independent. Thus we find

$$K_{31} = -2 \text{tr} \{ \tilde{\kappa}_{\xi} \Sigma^{-1} \tilde{\kappa}_{\xi} \} \text{tr} \{ M_{ba} M_{aa}^{-1} M_{ab} \}, \quad (A.31)$$
The term $K_{32}$ is of the order of a constant, and we can apply Lemma 15 and find the two contributions using the fact that $N_{au}$ and $N_{ay}$ are independent of $N_{by}$. We get that $E_{\xi_1}[N_{ya}N_{au}] = 0 = -n_a \hat{\kappa}_\xi$ (see (A.16)) and $T^{-1}E_{\xi_1}[M_{yb}M_{by}] = 0 = n_b \Sigma$ (see (A.20)), such that

$$K_{32} = -2T^{-1}E_{\xi_1}[\text{tr}\{\Sigma^{-1}M_{ya}M_{au}^{-1}M_{au}\hat{\kappa}_\xi\Sigma^{-1}M_{yb}M_{by}\}]$$

$$= -2T^{-1}\text{tr}\{\Sigma^{-1}E_{\xi_1}[N_{ya}N_{au}]\hat{\kappa}_\xi\Sigma^{-1}E_{\xi_1}[M_{yb}M_{by}]\}$$

$$= -2n_b n_a \text{tr}\{\Sigma^{-1}\hat{\theta}\hat{\kappa}_\xi\Sigma^{-1}\} = 2n_b n_a \text{tr}\{\hat{\kappa}_\xi\Sigma^{-1}\hat{\kappa}_\xi\}. \tag{A.32}$$

The term $K_{33}$ is zero because $N_{bu}$ and $N_{bu}$ are asymptotically independent of $N_{yb}$. The term $K_{34}$ is also zero because $(N_{ay}, N_{by})$ are asymptotically independent of $M_{yu}$; finally the term $K_{35}$ is zero because $(N_{au}, N_{by})$ is asymptotically independent of $N_{yy}$ (see Lemma 15).

Hence the contribution from $K_3$ is found from (A.32) and (A.31):

$$K_3 = 2(n_a n_b - \text{tr}\{M_{ba}M_{ua}^{-1}M_{ab}\})\text{tr}(\kappa_\xi\Sigma_{\omega_1z}^{-1}\kappa_\xi). \tag{A.33}$$

This completes the calculations, and it remains to compare (A.33) with (A.23) and (A.25)–(A.27), which gives

$$K_3 + K_4 + K_6 + K_7 = 0 = n_a n_b \text{tr}(\kappa_\xi\Sigma_{\omega_1z}^{-1}\kappa_\xi).$$

This completes the proof of Theorem 4.