EUROPEAN UNIVERSITY INSTITUTE

DEPARTMENT OF ECONOMICS

EUI Working Paper ECO No. 99/9

A Small Sample Correction for Tests of Hypotheses on the Cointegrating Vectors

Søren Johansen

BADIA FIESOLANA, SAN DOMENICO (FI)

Printed in Italy in March 1999
European University Institute
Badia Fiesolana
I-50016 San Domenico (FI)
Italy

A small sample correction for tests of hypotheses on the cointegrating vectors

Søren Johansen,
Economics Department,
European University Institute, Florence

February 24, 1999

Abstract

A correction factor, depending on sample size and parameters, is found for the likelihood ratio test for some linear hypotheses on the cointegrating space in a vector autoregressive model, where the adjustment coefficients are known. The main idea is to condition on the common trends when making inference on the cointegrating coefficients in order to calculate the Bartlett correction factor. Some simulation experiments illustrate the findings.

1 Introduction

We consider the $n-$dimensional vector autoregressive model for cointegration

$$\Delta X_t = \alpha\beta' X_{t-1} + \sum_{i=1}^{k-1} \Gamma_i \Delta X_{t-i} + \Phi d_t + \varepsilon_t,$$

where ε_t are i.i.d. $N_n(0, \Omega)$, with the purpose of finding a small sample correction for the likelihood ratio test for hypotheses on the cointegrating space spanned by β. It is known that the likelihood ratio test is χ^2 asymptotically distributed despite the fact that the asymptotic distribution of the estimator is mixed Gaussian; see Johansen (1988, 1996) and Ahn and Reinsel (1990). Many simulation studies indicate that there can be considerable size distortions when the χ^2 tables are used for inference, see for instance, Fachin (1997), Jacobsen and Gredenhoff (1998), and Jacobson, Vredin, and Warne (1998). We derive here a correction term to the likelihood ratio test statistic with the purpose of improving the approximation to the asymptotic χ^2 distribution. The correction is the so-called Bartlett correction, Bartlett (1937), and an overview of this type of correction can be found in Cribaro-Neto and Cordeiro (1996).

The actual distribution of the test statistic for β in the cointegrated vector autoregressive model depends on the sample size and on the many parameters of the model. Even though asymptotically this dependence vanishes, it is still important for finite T, and the simulations show that if the adjustment is slow the approximation can be very bad.

We derive here a correction factor that depends on sample size and the parameters in the model, such that it can be decided analytically when the approximation to the asymptotic χ^2 distribution is good and when a correction will improve the approximation. We face the usual problem with correction factors. If the factor is sufficiently close to one, we need not correct, and the asymptotic distribution is a good approximation, and if the factor is large, then the approximation by the asymptotic distribution is not very good, and the next order term may be needed. Hence the correction may not work. In between, there is an area where the correction factor is of moderate size and may be useful. Thus many simulations are needed to assess the usefulness of such a correction factor.

The problem to be solved is rather complicated and what follows is only a first attempt of a solution of a special case.

This paper is based upon the following ideas and observations

1. Since inference on β is asymptotically independent of inference on α the calculations will be done in the model where this parameter is fixed and known.

2. Since $\hat{\beta}$ is asymptotically mixed Gaussian and the discussion of asymptotic inference involves a conditioning argument on the asymptotic common trends, we condition throughout on the common trends when making inference.

In order to illustrate the conditioning idea consider the simple bivariate regression model in triangular form

$$Y_t = \beta X_t + \varepsilon_{1t},$$
$$\Delta X_t = \varepsilon_{2t},$$

where ε_t are i.i.d. $N_2(0, I_2)$. In this situation it is well known that

$$\hat{\beta} - \beta = \frac{\sum_{t=1}^T X_t \varepsilon_t}{\sum_{t=1}^T X_t^2},$$

has a mixed Gaussian distribution. The likelihood ratio test is equivalent to a test on

$$Q = \left(\hat{\beta} - \beta\right)^2 \sum_{t=1}^T X_t^2,$$

which is exactly $\chi^2(1)$, however. The reason for this is, that although the distribution of $\hat{\beta}$ has heavy tails and the value of the estimator calculated will often be extreme, this phenomenon is always followed by a small value of the information $\sum_{t=1}^T X_t^2$, such that the normalized contribution to the test statistic is not extreme. That is, the extreme observations in $\hat{\beta}$ are not extreme if measured by their "asymptotic conditional variance" $\left(\sum_{t=1}^T X_t^2\right)^{-1}$. A way of avoiding the issue of mixed Gaussian distribution is to consider conditional inference, that is, we consider the process X_t fixed and known.

Often, in the calculation of Bartlett corrections, one meets a factor of the form $s^{-2} = (T^{-1} \sum_{t=1}^T X_t^2)^{-1}$, for some process X_t. If X_t is a stationary process such that $s^2 \xrightarrow{P} E(s^2) = \sigma^2$, say, then we expand as follows

$$\frac{1}{s^2} = \frac{1}{\sigma^2 + (s^2 - \sigma^2)} = \frac{1}{\sigma^2} - \frac{1}{\sigma^2}(\frac{s^2}{\sigma^2} - 1) + \frac{1}{\sigma^2}(\frac{s^2}{\sigma^2} - 1)^2 - \dots$$

In this way negative powers can be replaced by positive powers, which greatly facilitates the calculations. In case the process X_t is a random walk, the limit $T^{-2} \sum_{t=1}^T X_t^2$ is not a constant and the above expansion does not help, since a random term appears in all numerators. Obviously, conditioning on X_t, will in this case fix all numerators and avoid the calculation of moments involving the random limit.

We try in the following to see how far this idea can take us in the cointegrated vector autoregressive model with known α. When the parameter α is fixed, we can transform the model into a cointegrated regression model, see (2) and (3), of the type considered by Phillips (1991).

In Section 2 we show that the autoregressive model with known adjustment coefficients reduces to an ordinary regression model but with non-stationary regressors. We derive likelihood ratio tests for hypotheses on the cointegrating space and normalize the regressors using the parameters from the true data generating process in order to be able to calculate the Bartlett correction. Section 3 contains the main result which gives the correction term for a general regression model and in Section 4 the results are applied to tests in the autoregressive model. Section 5 contains some simulation results which are used to illustrate the findings. The proof of the main result is given in an Appendix.

2 The autoregressive model and the test statistic

We show in this section that if α is known the model reduces to a regression model and we can formulate the test for a hypothesis on the cointegrating coefficients as a test on the coefficients to the non-stationary common stochastic trends. Since we condition on these the non-stationary regressors become fixed and deterministic, but they still satisfy some very special conditions, which are analogous to their stochastic properties when they are considered random.

We give the results for tests of the form $\beta = H\phi$, where $H(n \times s)$ is known and $\phi(s \times r)$ is a parameter to be estimated. We treat first the case where $s = r$, corresponding to a known cointegrating space, and then show how the general situation can be treated the same way.

The main result is the derivation of a general regression equation which contains all the above examples as special cases, and from which the likelihood ratio tests can easily be found. This implies that we can find the expansion of the likelihood ratio test which is needed for the calculation of the Bartlett correction.

2.1 The model equations

For notational reasons we focus here on the situation with two lags and mention when the results have to be modified for more lags. Consider the model

$$\Delta X_t = \alpha\beta' X_{t-1} + \Gamma_1 \Delta X_{t-1} + \Phi d_t + \varepsilon_t, \qquad t = 1, \ldots, T. \tag{1}$$

We assume that ε_t are i.i.d. $N_n(0, \Omega)$, and for the derivation of the estimator and likelihood ratio test we assume that the initial values X_0 and X_{-1} are fixed. We assume that the parameter $\alpha(n \times r)$ is known and that $\beta(n \times r)$, $\Gamma_1(n \times n)$, $\Omega(n \times n)$, and $\Phi(n \times d)$ are unknown and vary unrestrictedly. The deterministic terms d_t may contain for instance a constant, a linear term or seasonal dummies. The main property that we need here is that d_{t+1} is a linear function of d_t, ($d_{t+1} = Md_t$) a property that is satisfied in the above examples, and which implies that the processes $\beta' X_t$ and ΔX_t have expectations that are linear functions of d_t.

We exploit the fact that α is known to derive two equations from model (1). We let $\bar{\alpha}' = (\alpha'\alpha)^{-1}\alpha'$ and find

$$\bar{\alpha}'\Delta X_t = \beta' X_{t-1} + \bar{\alpha}'\Gamma_1 \Delta X_{t-1} + \bar{\alpha}'\Phi d_t + \bar{\alpha}'\varepsilon_t, \tag{2}$$

$$\alpha_\perp'\Delta X_t = \alpha_\perp'\Gamma_1 \Delta X_{t-1} + \alpha_\perp'\Phi d_t + \alpha_\perp'\varepsilon_t, \tag{3}$$

and finally, with $\omega = \bar{\alpha}'\Omega\alpha_\perp(\alpha_\perp'\Omega\alpha_\perp)^{-1}$, the conditional model of $\bar{\alpha}'\Delta X_t$ given $\alpha_\perp'\Delta X_t$ and the past is

$$\bar{\alpha}'\Delta X_t = \omega\alpha_\perp'\Delta X_t + \beta' X_{t-1} + \tilde{\Gamma}_1 \Delta X_{t-1} + \tilde{\Phi} d_t + \tilde{\varepsilon}_t, \tag{4}$$

where

$$\tilde{\Gamma}_1 = (\bar{\alpha}' - \omega\alpha_\perp')\Gamma_1,\ \tilde{\Phi} = (\bar{\alpha}' - \omega\alpha_\perp')\Phi,\ \tilde{\varepsilon}_t = (\bar{\alpha}' - \omega\alpha_\perp')\varepsilon_t.$$

Note that the errors are independent in (3) and (4). The errors $\alpha_\perp' \varepsilon_t$ are the permanent shocks which will be kept fixed in the calculation of the Bartlett correction, and the errors $(\bar{\alpha}' - \omega\alpha_\perp')\varepsilon_t' = (\alpha'\Omega\alpha)^{-1}\alpha'\Omega^{-1}\varepsilon_t$ are the transitory shocks which determine the random variation of the estimated β for given values of the common trends. The common trends in the process are $\alpha_\perp' \sum_{i=1}^{t} \varepsilon_i$ and the analysis of the distribution of the test statistic will be conducted conditional on these in the following.

Since the parameter β enters the conditional model (4) only, we derive the tests based on that equation.

2.2 Test of a simple hypothesis on the cointegrating space in the general model

We consider in this subsection the test that the cointegrating space is given, that is, $\beta = H\phi$ where ϕ is $(r \times r)$. It is of course very easy to derive the likelihood ratio test for a hypothesis on β in (4). We derive here an expression for the test statistic that depends on the parameters of the data generating process. This is convenient when we want to perform the very special calculations under a fixed probability measure specified by the parameters under the null hypothesis. These values we call the true values and denote them by $(\alpha^0, \beta^0 = H\phi^0, \Gamma_1^0, \Phi^0, \Omega^0)$, and indicate by a superscript ^{0} any parameter derived from these.

We assume that the process is $I(1)$, or equivalently that the roots of the characteristic polynomial are either greater than one in absolute value or equal to 1, and for $\Gamma^0 = I_n - \Gamma_1^0$, we assume that $\alpha_\perp^{0\prime}\Gamma^0\beta_\perp^0$ has full rank or equivalently again that $(\beta^0, \Gamma^{0\prime}\alpha_\perp^0)$ has full rank; see Johansen (1996). We define

$$C^0 = \beta_\perp^0(\alpha_\perp^{0\prime}\Gamma^0\beta_\perp^0)^{-1}\alpha_\perp^{0\prime}.$$

Note that

$$(I_n - C^0\Gamma^0)\beta_\perp^0 = 0,$$

such that $(I_n - C^0\Gamma^0)X_t = (I_n - C^0\Gamma^0)\bar{\beta}^0\beta^{0\prime}X_t$ is a linear function of $\beta^{0\prime}X_t$, and $C^0\Gamma^0X_t$ is a linear function of $\alpha_\perp^{0\prime}\Gamma^0X_t$. We therefore apply the decomposition

$$X_{t-1} = (I_n - C^0\Gamma^0)\bar{\beta}^0\beta^{0\prime}X_{t-1} + C^0\Gamma^0X_{t-1},$$

to decompose the process into a stationary and a non-stationary part. We define new parameters κ and δ by

$$\begin{aligned}\beta'X_{t-1} &= \beta'(I_n - C^0\Gamma^0)\bar{\beta}^0\beta^{0\prime}X_{t-1} + \beta'C^0\Gamma^0X_{t-1} \\ &= \kappa\beta^{0\prime}X_{t-1} + \delta\alpha_\perp^{0\prime}\Gamma^0X_{t-1}.\end{aligned}$$

Note that the $(r \times r)$ unknown parameters in $\kappa = \beta'(I_n - C^0\Gamma^0)\bar{\beta}^0$ are the adjustment coefficients to the known cointegration vectors in β^0, and that the $r \times (n-r)$ parameters $\delta = \beta'\beta_\perp^0(\alpha_\perp^{0\prime}\Gamma^0\beta_\perp^0)^{-1}$ are the coefficient to the non-stationary variables. Thus the regression equation (4) is

$$\bar{\alpha}^{0\prime}\Delta X_t = \omega\alpha_\perp^{0\prime}\Delta X_t + \kappa\beta^{0\prime}X_{t-1} + \delta\alpha_\perp^{0\prime}\Gamma^0X_{t-1} + \tilde{\Gamma}_1\Delta X_{t-1} + \tilde{\Phi}d_t + \tilde{\varepsilon}_t. \quad (5)$$

The hypothesis of a fixed cointegrating space spanned by β^0 is $H_0 : \delta = 0$, and the likelihood ratio test is derived by a regression of $\bar{\alpha}^{0\prime}\Delta X_t$ on $\alpha_\perp^{0\prime}\Gamma^0 X_{t-1}$ corrected for the variables $\alpha_\perp^{0\prime}\Delta X_t$, $\beta^{0\prime}X_{t-1}$, ΔX_{t-1}, and d_t.

From (3) we find that in the regression (5) we can replace $\alpha_\perp^{0\prime}\Delta X_t$ by $\alpha_\perp^{0\prime}\varepsilon_t$ since we are correcting for ΔX_{t-1} and d_t in the regression anyway, and by summation of (3) we find

$$\alpha_\perp^{0\prime}\Gamma^0 X_{t-1} = -\alpha_\perp^{0\prime}\Gamma_1^0\Delta X_{t-1} + \alpha_\perp^{0\prime}\sum_{i=1}^{t-1}(\varepsilon_i + \Phi^0 d_i) + \alpha_\perp^{0\prime}(X_0 - \Gamma_1^0 X_{-1}). \qquad (6)$$

This shows that in the regression (5) we can replace $\alpha_\perp^{0\prime}\Gamma^0 X_{t-1}$ by $a_{t-1}^0 = \alpha_\perp^{0\prime}\sum_{i=1}^{t-1}(\varepsilon_i + \Phi^0 d_i) + \alpha_\perp^{0\prime}(X_0 - \Gamma_1^0 X_{-1})$ since we are correcting for ΔX_{t-1} in the regression. Note that if d_t contains a constant we can drop the initial values in the definition of a_{t-1}^0. In the calculations in Section 3 we keep $\alpha_\perp^{0\prime}\varepsilon$ fixed (and hence a_{t-1}^0) and we define the normalized deterministic regressors

$$b_t = (\alpha_\perp^{0\prime}\Omega^0\alpha_\perp^0)^{-\frac{1}{2}}\alpha_\perp^{0\prime}\varepsilon_t, \qquad (7)$$

and let a_{t-1} be a_{t-1}^0 but orthogonalized to d_t and b_t, and moreover normalized, such that

$$\sum_{t=1}^T a_{t-1}d_t' = 0, \ \sum_{t=1}^T a_{t-1}b_t' = 0, \ \sum_{t=1}^T a_{t-1}a_{t-1}' = I_{n-r}.$$

Similarly we define the normalized errors as

$$U_t = (\alpha^{0\prime}\Omega^{0-1}\alpha^0)^{-\frac{1}{2}}\alpha^{0\prime}\Omega^{0-1}\varepsilon_t = (\alpha^{0\prime}\Omega^{0-1}\alpha^0)^{-\frac{1}{2}}(\bar{\alpha}' - \omega^0\alpha_\perp')\varepsilon_t. \qquad (8)$$

We define the unconditional variance

$$\Sigma^0 = \mathrm{Var}\begin{pmatrix} \beta^{0\prime}X_t \\ \Delta X_t \end{pmatrix}.$$

From the Granger representation of X_t, see Johansen (1996),

$$X_t = C^0\sum_{i=1}^t(\varepsilon_i + \Phi^0 d_i) + \sum_{i=0}^\infty C_i^0(\varepsilon_{t-i} + \Phi^0 d_{t-i}), \qquad (9)$$

we find with $d_{t-i} = M^{-i}d_t$, that

$$E(\beta' X_t) = \textstyle\sum_{i=0}^\infty C_i^0\Phi^0 d_{t-i} = (\sum_{i=0}^\infty C_i^0\Phi^0 M^{-i})d_t$$
$$E(\Delta X_t) = \left(C^0\Phi^0 + \textstyle\sum_{i=0}^\infty C_i^0\Phi^0 M^{-i}(I - M^{-1})\right)d_t$$

Thus the mean $E(X_t'\beta^0, \Delta X_t')$ is a linear function of the deterministic terms. We define the normalized process

$$Y_t = \Sigma^{0-\frac{1}{2}}\begin{pmatrix} \beta^{0\prime}X_t - E(\beta^{0\prime}X_t) \\ \Delta X_t - E(\Delta X_t) \end{pmatrix}. \qquad (10)$$

If there are more than two lags in the model we have to define the process Y_t by stacking $\beta^{0\prime}X_t, \Delta X_t, \ldots, \Delta X_{t-k+2}$, which becomes of dimension $m = r + (k-1)n$.

The regression equation (5) can then be written in the form

$$\underset{(r)}{Z_t} = \xi_1 \underset{(n-r)}{a_{t-1}} + \xi_2 \underset{(m)}{Y_{t-1}} + \xi_3 \underset{(n-r)}{b_t} + \xi_4 \underset{(d)}{d_t} + \xi_5 \underset{(r)}{U_t}, \tag{11}$$

with $Z_t = \bar{\alpha}'\Delta X_t$, and

$$Y_t = \Sigma^{0-\frac{1}{2}}(X_t'\beta^0 - E(X_t'\beta^0),\ \Delta X_t' - E(\Delta X_t'), \dots,\ \Delta X_{t-k+2}' - E(\Delta X_{t-k+2}'))' \tag{12}$$

and suitable coefficients $\xi_1, \dots, \xi_5$. Note that centering Y_{t-1} to have mean zero introduces a correction to the coefficient of d_t since the mean of $\beta' X_t$ and ΔX_t are linear functions of d_t. The test that the cointegrating space is sp(β^0) is the test that $\xi_1 = 0$. Below the variables are indicated their dimensions. The maximum likelihood estimator or regression estimator for ξ_1 satisfies

$$\xi_5 S_{ua.y,b,d} = (\hat{\xi}_1 - \xi_1) S_{aa.y,b,d},$$

where we have used the notation for the product moments of $(U_t, Y_{t-1}, a_{t-1}, b_t, d_t)$:

$$\sum_{t=1}^{T} \begin{pmatrix} U_t \\ Y_{t-1} \\ a_{t-1} \\ b_t \\ d_t \end{pmatrix} \begin{pmatrix} U_t \\ Y_{t-1} \\ a_{t-1} \\ b_t \\ d_t \end{pmatrix}' = \begin{pmatrix} S_{uu} & S_{uy} & S_{ua} & S_{ub} & S_{ud} \\ S_{yu} & S_{yy} & S_{ya} & S_{yb} & S_{yd} \\ S_{au} & S_{ay} & I_{n-r} & 0 & 0 \\ S_{bu} & S_{by} & 0 & S_{bb} & S_{bd} \\ S_{du} & S_{dy} & 0 & S_{db} & S_{dd} \end{pmatrix}.$$

Let $V_t, W_t,$ and F_t be any of these processes then we also need the conditional product moments

$$S_{vw.f} = S_{vw} - S_{vf} S_{ff}^{-1} S_{fw}.$$

With this notation we find the likelihood ratio test of the hypothesis that the cointegrating space is given by sp$(\beta^0) =$sp(H) can be tested as $\xi_1 = 0$ by the likelihood ratio test

$$-2 \log LR(\beta = H\phi) = -T \log \frac{|S_{uu.y,a,b,d}|}{|S_{uu.y,b,d}|}. \tag{13}$$

2.3 Test for linear restrictions on the cointegrating space

We consider in this section the likelihood ratio test for the hypothesis $\beta = H\phi$, for $H(n \times s)$ and $\phi(s \times r)$ $(r \leq s \leq n)$. We let $L(\beta, \Phi, \Gamma_1, \Omega)$ denote the Gaussian likelihood and define the concentrated likelihood

$$L(\beta) = \max_{\Gamma_1, \Omega, \Phi} L(\beta, \Phi, \Gamma_1, \Omega)$$

and use the multiplicative property of likelihood ratio test to see that for $\beta^0 = H\phi^0$,

$$\frac{\max_{\beta=H\phi} L(\beta)}{\max L(\beta)} = \frac{L(\beta^0)}{\max L(\beta)} \Big/ \frac{L(\beta^0)}{\max_{\beta=H\phi} L(\beta)}$$

and hence

$$-2\log LR(\beta = H\phi) = \ 2\log LR(\beta = \beta^0|\beta = H\phi) - 2\log LR(\beta = \beta^0). \tag{14}$$

The second term was found in (13) and we investigate here the test of a simple hypothesis under the assumption that $\beta = H\phi$ and find that (11) still holds with a small modification of the dimensions.

Consider equation (2), in particular the term

$$\beta' X_{t-1} = \phi' H' X_{t-1} = \phi' H' C^0\Gamma^0 X_{t-1} + \kappa\beta^{0\prime} X_{t-1},$$

where $\kappa = \phi' H'(I_n - C^0\Gamma^0)\bar{\beta}^0(r\times r)$. If $\beta^0 = H\phi^0$, is the true value of the parameter, then $\beta^0_\perp = (H_\perp, \bar{H}\phi^0_\perp)$, such that

$$\begin{aligned}\phi' H' C^0\Gamma^0 X_{t-1} &= \phi' H'(H_\perp, \bar{H}\phi^0_\perp)(\alpha'_\perp\Gamma^0\beta^0_\perp)^{-1}\alpha'_\perp\Gamma^0 X_{t-1}\\ &= \phi'\phi^0_\perp(0_{(s-r)\times(n-s)}, I_{s-r})(\alpha'_\perp\Gamma^0\beta^0_\perp)^{-1}\alpha'_\perp\Gamma^0 X_{t-1}\\ &= \delta A\alpha'_\perp\Gamma^0 X_{t-1},\end{aligned}$$

where $\delta = \phi'\phi^0_\perp$ of dimension $r\times(s-r)$ is a parameter and

$$A = (0_{(s-r)\times(n-s)}, I_{s-r})(\alpha'_\perp\Gamma^0\beta^0_\perp)^{-1}$$

is a known $(s-r)\times(n-r)$ matrix, such that the number of common trends that enter the equation is only $s-r$ not $n-r$ as in the previous section. Thus the regression equation (5) still holds with $\alpha'_\perp\Gamma^0 X_{t-1}$ replaced by $A\alpha'_\perp\Gamma^0 X_{t-1}$. We can, as before, replace $\alpha'_\perp\Delta X_t$ with $\alpha'_\perp\varepsilon_t$ and $\alpha'_\perp\Gamma^0 X_{t-1}$ by $\alpha'_\perp\sum_{i=1}^t(\varepsilon_i + \Phi d_i)$, possibly with initial values, but the matrix A reduces the number of trends, so we define in this case $a_{t-1} = A\alpha'_\perp\sum_{i=1}^t(\varepsilon_i + \Phi d_i)$ corrected for b_t and d_t and normalized such that $\sum_{t=1}^T a_{t-1}a'_{t-1} = I_{s-r}$. Thus equation (11) holds but with the dimension of a reduced to $s-r$.

The test for a simple hypothesis on β, when we assume that $\beta = H\phi$, can be tested as $\xi_1 = 0$ and gives the statistic (13) only with a of dimension $s-r$, and (14) gives an expression for $-2\log LR(\beta = H\phi)$.

2.4 An expansion of the likelihood ratio test

In order to cover the two examples in a general equation, and in order to apply the results in the further analyses of the Bartlett correction, we formulate the main result for the equation:

$$\underset{(n_u)}{Z_t} = \xi_1 \underset{(n_a)}{a_{t-1}} + \xi_2 \underset{(n_y)}{Y_{t-1}} + \xi_3 \underset{(n_b)}{b_t} + \xi_4 \underset{(n_d)}{d_t} + \xi_5 \underset{(n_u)}{U_t}. \tag{15}$$

This formulation covers the cases discussed in subsections 2.2 and 2.3 by suitable choices of Z_t, a_t, Y_t, b_t, d_t, and U_t. We assume that Z_t is a linear function of ΔX_t, Y_t is the stacked process $(\beta' X_{t-1}, \Delta X_{t-1}, \ldots, \Delta X_{t-k+2})$ normalized to have mean zero and variance I_{n_y}, a_{t-1} is a linear function of $\alpha'_\perp\sum_{i=1}^{t-1}\varepsilon_i$ possibly modified by deterministic terms and b_t is a linear function of $\alpha'_\perp\varepsilon_t$. Finally $d_{t+1} = Md_t$.

We derive an expansion for the likelihood ratio test that $\xi_1 = 0$. We find from (13)

$$\begin{aligned}
-2\log LR(\xi_1 = 0) &= -T \log \tfrac{|S_{uu.y,a,b,d}|}{|S_{uu.y,b,d}|} \\
&= -T \log |I_{n_u} - T^{-1}(T^{-1}S_{uu.y,b,d})^{-1} S_{ua.y,b,d} S_{aa.y,b,d}^{-1} S_{au.y,b,d}| \\
&= -T \log |I_{n_u} - T^{-1}Q| \stackrel{1}{=} tr\{Q\} + \tfrac{1}{2T} tr\{Q^2\},
\end{aligned}$$

where the notation $\stackrel{s}{=}$ indicates that we have kept terms of the order T^{-s}, $s = 0, 1, \dots$, and

$$Q = (T^{-1}S_{uu.y,b,d})^{-1} S_{ua.y,b,d} S_{aa.y,b,d}^{-1} S_{au.y,b,d} \in O_P(1). \tag{16}$$

In order to simplify the expression for Q we note that since a_{t-1} is orthogonal to b_t and d_t, we get

$$\begin{aligned}
S_{aa.y,b,d} &= S_{aa.b,d} - S_{ay.b,d} S_{yy.b,d}^{-1} S_{ya.b,d} \\
&= S_{aa} - S_{ay} S_{yy.b,d}^{-1} S_{ya} = I_{n_u} - T^{-1}B \qquad (17)\\
S_{au.y,b,d} &= S_{au.b,d} - S_{ay.b,d} S_{yy.b,d}^{-1} S_{yu.b,d} \\
&= S_{au} - S_{ay} S_{yy.b,d}^{-1} S_{yu.b,d} = N - T^{-\frac{1}{2}}A, \qquad (18)
\end{aligned}$$

where

$$\begin{aligned}
N &= S_{au} \qquad (19)\\
B &= S_{ay}(T^{-1}S_{yy.b,d})^{-1} S_{ya}, \qquad (20)\\
A &= S_{ay}(T^{-1}S_{yy.b,d})^{-1} (T^{-\frac{1}{2}} S_{yu.b,d}), \qquad (21)
\end{aligned}$$

are $O_P(1)$ and N is distributed as $N_{n_a \times n_u}(0, I_{n_a} \otimes I_{n_u})$, when we condition on $\alpha_\perp' \varepsilon$.

We find

$$\begin{aligned}
& T^{-1}S_{uu.y,b,d} = T^{-1}S_{uu.b,d} - T^{-1}S_{uy.b,d} S_{yy.b,d}^{-1} S_{yu.b,d} \\
= & \; T^{-1}S_{uu} - T^{-1}S_{ud}S_{dd}^{-1}S_{du} - T^{-1}S_{ub.d}S_{bb.d}^{-1}S_{bu.d} - T^{-1}S_{uy.b,d}S_{yy.b,d}^{-1}S_{yu.b,d} \\
= & \; I_{n_u} + (T^{-1}S_{uu} - I_{n_u}) - T^{-1}(S_{ud}S_{dd}^{-1}S_{du} + S_{ub.d}S_{bb.d}^{-1}S_{bu.d} + S_{uy.b,d}S_{yy.b,d}^{-1}S_{yu.b,d}) \\
= & \; I_{n_u} + T^{-\frac{1}{2}}D_1 - T^{-1}D_2. \qquad (22)
\end{aligned}$$

We then find from (16), (17), and (18) that

$$\begin{aligned}
& tr\{Q\} \\
= & \; tr\{(I_{n_u} + T^{-\frac{1}{2}}D_1 - T^{-1}D_2)^{-1}(N - T^{-\frac{1}{2}}A)'(I_{n_a} - T^{-1}B)^{-1}(N - T^{-\frac{1}{2}}A)\} \\
\stackrel{1}{=} & \; tr\{N'N - T^{-\frac{1}{2}}(N'A + A'N + D_1N'N) \\
+ & \; T^{-1}(D_2N'N + D_1^2N'N + D_1(N'A + A'N) + N'BN + A'A)\},
\end{aligned}$$

where we have kept terms of order T^{-1}. The notation has been chosen such that the power of T in front indicates the order of the various terms.

Similarly we find

$$tr\{Q^2\} \stackrel{0}{=} tr\{(N'N)^2\}.$$

The asymptotic properties in the usual analysis, where ε is considered random, are given in terms of a Brownian motion $W(u)$:

$$T^{-\frac{1}{2}}\sum_{i=1}^{[Tu]}\varepsilon_i \xrightarrow{w} W(u).$$

We define the independent standard Brownian motions $B_1(u) = (\alpha_\perp'\Omega\alpha_\perp)^{-\frac{1}{2}}\alpha_\perp' W(u)$ and $B_2(u) = (\alpha'\Omega^{-1}\alpha)^{-\frac{1}{2}}\alpha'\Omega^{-1}W(u)$.

The asymptotic distribution of $-2\log LR$, for the hypothesis that the cointegrating space is spanned by β, is given by

$$-2\log LR \overset{0}{=} tr\{Q\} \xrightarrow{w} tr\{\int_0^1 (dB_2)F' \left(\int_0^1 FF'du\right)^{-1}\int_0^1 F(dB_2)'\}, \tag{23}$$

where F is constructed from B_1 depending on the deterministic terms. If $d_t = 0$, then $F = B_1$, and if for instance $\Phi d_t = \mu_0$, then $\alpha_\perp'\sum_{i=1}^{t-1}\varepsilon_i + \alpha_\perp'\mu_0(t-1)$ is corrected for the mean. The process is linear in one direction, $\alpha_\perp'\mu_0$, and a random walk in $n-r-1$ directions such that

$$\begin{array}{l} F_i(u) = B_i(u) - \int_0^1 B_i(s)ds, i = 1, \dots, n-r-1, \\ F_{n-r}(u) = u - \frac{1}{2}. \end{array} \tag{24}$$

For fixed value of F the distribution of $\int_0^1 F(dB_2)'$ is Gaussian with mean zero and variance matrix $\int_0^1 FF'du \otimes I_{n_u}$ such that the limit distribution of Q for fixed F is χ^2 with degrees of freedom $n_u n_a$. Since this distribution does not depend on the conditioning random variable, the marginal distribution is also $\chi^2(n_u n_a)$.

The idea in the following is to take the consequence of this conditioning argument, and condition already from the beginning on the common trends $\alpha_\perp'\sum_{i=1}^t \varepsilon_i$, which give rise to the Brownian motion B_1 and F. The paper by Rothenberg (1988) contains another conditioning idea, due to Cavanagh (1983), which in the present context would mean conditioning on the zero order term $N'N$ and work out the moments of the remaining terms in this conditional distribution. This will not be attempted here.

Bartlett (1937) proposed to improve the approximation to the limit distribution of the likelihood ratio test by adjusting the finite sample distribution to have the same mean as the limit distribution. This simple correction turns out in a number of cases to have a good effect on the approximation, and we shall therefore try to calculate the expectation $E[-2\log LR|\alpha_\perp'\varepsilon]$. The exact calculation is not possible but we can derive an expression for the first order approximation. It turns out that this approximation depends on the parameters and thus will have to be estimated in practice, which gives some extra uncertainty.

Thus we want to calculate

$$E[-2\log LR|\alpha_\perp'\varepsilon] \overset{1}{=} E[tr\{Q\}|\alpha_\perp'\varepsilon] + \frac{1}{2T}E[tr\{Q^2\}|\alpha_\perp'\varepsilon], \tag{25}$$

where

$$\begin{array}{l} E[tr\{Q\}|\alpha_\perp'\varepsilon] \\ \overset{1}{=} tr\{E[N'N|\alpha_\perp'\varepsilon]\} - T^{-\frac{1}{2}}tr\{E[(2A'N + D_1N'N)|\alpha_\perp'\varepsilon]\} \\ +T^{-1}tr\{E[(D_2N'N + D_1^2N'N + 2D_1N'A + N'BN + A'A)|\alpha_\perp'\varepsilon]\}, \end{array} \tag{26}$$

and

$$E[tr\{Q^2\}|\alpha_\perp'\varepsilon] \overset{0}{=} tr\{E[(N'N)^2|\alpha_\perp'\varepsilon]\}. \tag{27}$$

2.5 The fixed regressors

We insert a few comments on the fixed regressors a_{t-1} and b_t, defined in terms of $\alpha_\perp' \sum_{i=1}^{t-1} \varepsilon_i$, and $\alpha_\perp' \varepsilon_t$.

By the normalization chosen above we have

$$S_{ad} = \sum_{t=1}^{T} a_{t-1} d_t' = 0,\ S_{ab} = \sum_{t=1}^{T} a_{t-1} b_t' = 0,\ S_{aa} = \sum_{t=1}^{T} a_{t-1} a_{t-1}' = I_{n_a}.$$

When we do not condition on $\alpha_\perp' \sum_{i=1}^{t-1} \varepsilon_i$ we have the following relations

$$S_{bd} S_{dd}^{-1} S_{db} \in O_P(1), \tag{28}$$

$$\sum_{t=1}^{T} a_{t-1} a_{t-1-k}' \overset{P}{\to} I_{n_a},\ \text{for all } k, \tag{29}$$

$$T^{-1} \sum_{t=1}^{T} b_t b_{t-k}' \overset{P}{\to} \begin{cases} I_{n_b} & \text{if } k = 0, \\ 0 & \text{if } k \neq 0, \end{cases} \tag{30}$$

$$\sum_{t=1}^{T} b_{t-k} a_t' - \sum_{t=1}^{T} b_t a_t' \overset{P}{\to} 0 \ \text{ if } k \geq 0, \tag{31}$$

Finally we have that $\sum_{t=1}^{T} b_t a_t'$ converges weakly to a limit $\mathcal{I}_{ba}$. The limit depends on the deterministic terms, such that if $d_t = 0$, then

$$\sum_{t=1}^{T} b_t a_t' \overset{w}{\to} (I_{n_b} + \int_0^1 (dB_2) B_1') \left(\int_0^1 B_1 B_1' du \right)^{-\frac{1}{2}} = \mathcal{I}_{ba}, \tag{32}$$

whereas if $\Phi d_t = \mu_0$ then

$$\sum_{t=1}^{T} b_t a_t' \overset{w}{\to} ((I_{n_b-1}, 0_{(n_b-1)\times 1})' + \int_0^1 (dB_2) F') \left(\int_0^1 F F' du \right)^{-\frac{1}{2}}, \tag{33}$$

with F given by (24).

When conditioning on the sequence $\alpha_\perp' \varepsilon$ we assume that such relations hold for the sequence we are fixing. That is, we shall replace $\sum_{t=1}^{T} a_{t-1} a_{t-1-k}'$ by $\sum_{t=1}^{T} a_{t-1} a_{t-1}' = I_{n_a}$, $T^{-1} \sum_{t=1}^{T} b_t b_{t-k}'$ by I_{n-r} or 0, and, for $k \geq 0$, $\sum_{t=1}^{T} b_{t-k} a_t'$ by the matrix $\mathcal{I}_{ba}$ in the limit results. It turns out that although $\mathcal{I}_{ba}$ enters some of the intermediate results, the final result does not involve $\mathcal{I}_{ab}$, such that there is no need to calculate the expectation when it appears.

The papers by Kiviet and Phillips (1997a,1997b) deal with a regression model which resembles (15):

$$y_t = \lambda y_{t-1} + \gamma' x_t + \varepsilon_t,$$

and derive approximations for the bias of the least squares estimator of (λ, γ). The calculations are very similar to those needed here, but the scope differs in a number of aspects. We are analysing a multivariate situation rather than a univariate, and want an approximation to the likelihood ratio test statistic rather than the bias. We are interested only in the first order term, which implies that many of the formulae look simpler, and involve smaller matrices, since the above relations for the regressors a_{t-1} and b_t allow some simplifications of the general terms.

3 The main results on the Bartlett correction

We introduce here the coefficients that are used to express the main result in Proposition 1, which gives an approximation to the various terms entering the expression for the conditional expectation of the likelihood ratio test of the hypothesis $\xi_1 = 0$ in (15).

In the following we assume that the parameters $(\alpha, \beta, \Gamma_1, \ldots, \Gamma_{k-1}, \Phi, \Omega)$ are so chosen that the process is $I(1)$. Thus the non-stationary part of the process is started in the initial values, but the stationary part of the process, $\beta' X_t - E(\beta' X_t)$ and $\Delta X_t - E(\Delta X_t)$, are given their invariant distribution. We do not use the notation with a superscript ^{0} since we need not distinguish between the parameter value and the true value in this section.

This stationarity of $\beta' X_t - E(\beta' X_t)$ and $\Delta X_t - E(\Delta X_t)$ of course holds when we do not condition on $\alpha'_\perp \varepsilon$, and then Y_t, see (12), has the representation

$$Y_t = \sum_{\nu=0}^{\infty} C_\nu \varepsilon_{t-\nu} = \sum_{\nu=0}^{\infty} (\theta_\nu U_{t-\nu} + \psi_\nu b_{t-\nu}), \tag{34}$$

see (7) and (8) for suitable matrices θ_ν, ψ_ν, see (47). When we condition on $\alpha'_\perp \varepsilon_t$, however, then Y_t is no longer stationary, since its mean depends on t.

Note that the unconditional autocorrelation function is

$$\text{Cov}(Y_t, Y_{t+h}) = \gamma(h) = \sum_{\nu=0}^{\infty} (\theta_\nu \theta'_{\nu+h} + \psi_\nu \psi'_{\nu+h}),\ \gamma(0) = I_{n_y}. \tag{35}$$

We define the coefficients

$$\theta = \sum_{\nu=0}^{\infty} \theta_\nu, \psi = \sum_{\nu=0}^{\infty} \psi_\nu, \tag{36}$$

such that long-run variance of the stationary process Y_t is $\theta\theta' + \psi\psi'$, whereas the long-run variance of Y_t conditional on the common trends is $\theta\theta'$.

We can now formulate the main result on the test that $\xi_1 = 0$ in the regression (15). The result is given as a first order approximation of $E[-2 \log LR | \alpha'_\perp \varepsilon]$ expressed in terms of the coefficients θ_ν, ψ_ν, and $\gamma(h)$ and the dimensions n_u, n_a, n_b, n, n_d, and n_y, and M. First, however, we give a technical result which is the basis for the various applications later.

Proposition 1 *In the regression equation (15) we have the results*

$$tr\{E[(N'N)^2|\alpha'_\perp\varepsilon]\} = n_a n_u(n_a+n_u+1) \quad (37)$$
$$tr\{E[N'N|\alpha'_\perp\varepsilon]\} = n_a n_u \quad (38)$$

$$\begin{aligned}&T^{\frac{1}{2}}tr\{E[AN'|\alpha'_\perp\varepsilon]\}\\ &\to n_a(n_u+1)tr\{\theta\theta'\}+n_u tr\{\mathcal{I}_{ba}\mathcal{I}_{ab}\psi'\psi\}-n_a\textstyle\sum_{\nu=0}^{\infty}tr\{M^{\nu+1}\}tr\{\theta'_\nu\theta\}\\ &-n_a\textstyle\sum_{\nu=0}^{\infty}[tr\{\gamma(\nu+1)\theta\theta'_\nu\}+tr\{\gamma(\nu+1)\}tr\{\theta\theta'_\nu\}]\end{aligned} \quad (39)$$

$$tr\{E[A'A|\alpha'_\perp\varepsilon]\} \to n_a n_u tr\{\theta\theta'\}+n_u tr\{\mathcal{I}_{ba}\mathcal{I}_{ab}\psi'\psi\} \quad (40)$$

$$tr\{E[N'BN|\alpha'_\perp\varepsilon]\} \to n_a(n_a+n_u+1)tr\{\theta\theta'\}+n_u tr\{\mathcal{I}_{ba}\mathcal{I}_{ab}\psi'\psi\} \quad (41)$$

$$T^{\frac{1}{2}}tr\{E[D_1N'N|\alpha'_\perp\varepsilon]\} \to n_a n_u(n_u+1) \quad (42)$$
$$tr\{E[(D_2N'N|\alpha'_\perp\varepsilon]\} \to n_a n_u(n_d+n_b+n_y) \quad (43)$$
$$tr\{E[D_1N'A|\alpha'_\perp\varepsilon]\} \to 0 \quad (44)$$
$$tr\{E[D_1^2N'N|\alpha'_\perp\varepsilon]\} \to n_a n_u(n_u+1) \quad (45)$$

where the parameters θ_ν, ψ_ν, θ, ψ, *and* $\gamma(h)$ *are given in (34), (35), and (36). The definition of* (N, A, B, D_1, D_2) *are given in (19), (20), (21), and (22).*

Note that the correction terms depend on the conditioning variable through the quantity $\mathcal{I}_{ab}$, see (32) and (33). The proof of this result is given in the Appendix.

Proposition 2 *The first order approximation to the conditional expectation of the likelihood ratio test statistic for* $\xi_1 = 0$ *in (15) is given by*

$$\begin{aligned}&E[-2\log LR|\alpha'_\perp\varepsilon]\\ &\overset{1}{=} n_u n_a+\frac{n_u n_a}{T}[\frac{1}{2}(n_u+n_a+1)+(n_d+n_b+n_y)]+\frac{n_a}{T}[(n_a-1)v+2(c+c_d)].\end{aligned}$$

with

$$\begin{aligned}v &= tr\{\theta\theta'\}\\ c_d &= \sum_{\nu=0}^{\infty}tr\{M^{\nu+1}\}tr\{\theta\theta'_\nu\}\\ c &= \sum_{\nu=0}^{\infty}(tr\{\gamma(\nu+1)\theta\theta'_\nu\}+tr\{\gamma(\nu+1)\}tr\{\theta\theta'_\nu\})\end{aligned}$$

The proof follows from Proposition 1 by using (25), (26), and (27). Note that the correction term to the likelihood ratio test statistic does not depend on the conditioning random variables in the sense that $\mathcal{I}_{ab}$ is no longer present. In the following we give more explicit formulae using the properties of the underlying process.

The process Y_t, see (10) and (34), satisfies the equation

$$\Sigma^{\frac{1}{2}} Y_t = P \Sigma^{\frac{1}{2}} Y_{t-1} + B\varepsilon_t,$$

where

$$P(n_y \times n_y) = \begin{pmatrix} I_r + \beta'\alpha & \beta'\Gamma_1 & \cdots & \beta'\Gamma_{k-2} & \beta'\Gamma_{k-1} \\ \alpha & \Gamma_1 & \cdots & \Gamma_{k-2} & \Gamma_{k-1} \\ 0 & I_n & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & I_n & 0 \end{pmatrix}, B(n_y \times n) = \begin{pmatrix} \beta' \\ I_n \\ \vdots \\ 0 \\ 0 \end{pmatrix}. \quad (46)$$

We find the representation

$$\begin{aligned} Y_t &= \sum_{\nu=0}^{\infty} \Sigma^{-\frac{1}{2}} P^{\nu} B \varepsilon_{t-\nu} \qquad (47) \\ &= \sum_{\nu=0}^{\infty} \Sigma^{-\frac{1}{2}} P^{\nu} B\alpha(\alpha'\Omega^{-1}\alpha)^{-\frac{1}{2}} U_{t-\nu} + \Sigma^{-\frac{1}{2}} P^{\nu} B\Omega\alpha_{\perp}(\alpha_{\perp}'\Omega\alpha_{\perp})^{-\frac{1}{2}} b_{t-\nu} \end{aligned}$$

and use it to calculate the covariance matrix Σ

$$\begin{aligned} vec(\Sigma) &= vec(\sum_{\nu=0}^{\infty} P^{\nu} B\Omega B' P'^{\nu}) = \sum_{\nu=0}^{\infty} (P \otimes P)^{\nu} vec\,(B\Omega B') \\ &= (I_{n_y} \otimes I_{n_y} - P \otimes P)^{-1} vec\,(B\Omega B')\,. \end{aligned}$$

Comparing (47) with (34) we find

$$\theta_{\nu} = \Sigma^{-\frac{1}{2}} P^{\nu} B\alpha(\alpha'\Omega^{-1}\alpha)^{-\frac{1}{2}},\ \psi_{\nu} = \Sigma^{-\frac{1}{2}} P^{\nu} B\Omega\alpha_{\perp}(\alpha_{\perp}'\Omega\alpha_{\perp})^{-\frac{1}{2}}.$$

Note that by the definition of Σ

$$\sum_{\nu=0}^{\infty} (\theta_{\nu}\theta_{\nu}' + \psi_{\nu}\psi_{\nu}') = \mathrm{Var}(Y_t) = I_{n_y}. \qquad (48)$$

3.1 Calculation of $v = tr\{\theta\theta'\}$

We first calculate the conditional long-run variance of Y_t as given by $\theta\theta'$. We find

$$\theta = \sum_{\nu=0}^{\infty} \theta_{\nu} = \sum_{\nu=0}^{\infty} \Sigma^{-\frac{1}{2}} P^{\nu} B\alpha(\alpha'\Omega^{-1}\alpha)^{-\frac{1}{2}} = \Sigma^{-\frac{1}{2}} (I_{n_y} - P)^{-1} B\alpha(\alpha'\Omega^{-1}\alpha)^{-\frac{1}{2}}, \qquad (49)$$

and hence

$$\theta\theta' = \Sigma^{-\frac{1}{2}} (I_{n_y} - P)^{-1} B\alpha(\alpha'\Omega^{-1}\alpha)^{-1}\alpha' B' (I_{n_y} - P')^{-1} \Sigma^{-\frac{1}{2}}. \qquad (50)$$

Since

$$(I_{n_y} - P)(I_{n_u}, 0, \ldots, 0)' = -B\alpha$$

we find

$$\theta\theta' = \Sigma^{-\frac{1}{2}} \begin{pmatrix} (\alpha'\Omega^{-1}\alpha)^{-1} & 0 \\ 0 & 0 \end{pmatrix} \Sigma^{-\frac{1}{2}}.$$

We define the coefficient matrix

$$V = \Sigma^{\frac{1}{2}}\theta\theta'\Sigma^{-\frac{1}{2}} = \begin{pmatrix} (\alpha'\Omega^{-1}\alpha)^{-1} & 0 \\ 0 & 0 \end{pmatrix} \Sigma^{-1}. \quad (51)$$

Note that

$$tr\{\theta'\theta\} = tr\{\begin{pmatrix} I_{n_u} \\ 0 \end{pmatrix}' \Sigma^{-1} \begin{pmatrix} I_{n_u} \\ 0 \end{pmatrix} (\alpha'\Omega^{-1}\alpha)^{-1}\} = tr\{\Sigma_{\beta\beta}^{-1}(\alpha'\Omega^{-1}\alpha)^{-1}\},$$

where

$$\Sigma_{\beta\beta} = Var(\beta' X_t|\Delta X_t, \Delta X_{t-1}, \dots, \Delta X_{t-k+2}).$$

3.2 Calculation of c and c_d

We note that

$$\theta'_{\eta+\nu+1} = \theta'_\eta(\Sigma^{\frac{1}{2}}P'^{\nu+1}\Sigma^{-\frac{1}{2}}),\ \psi'_{\eta+\nu+1} = \psi'_\eta(\Sigma^{\frac{1}{2}}P'^{\nu+1}\Sigma^{-\frac{1}{2}}),$$

such that

$$\begin{aligned} \gamma(\nu+1) &= \sum_{\eta=0}^{\infty}(\theta_\eta\theta'_{\eta+\nu+1} + \psi_\eta\psi'_{\eta+\nu+1}) \\ &= \sum_{\eta=0}^{\infty}(\theta_\eta\theta'_\eta + \psi_\eta\psi'_\eta)(\Sigma^{\frac{1}{2}}P'^{\nu+1}\Sigma^{-\frac{1}{2}}) \\ &= \Sigma^{\frac{1}{2}}P'^{\nu+1}\Sigma^{-\frac{1}{2}}, \end{aligned}$$

since by (35)

$$\sum_{\eta=0}^{\infty}(\theta_\eta\theta'_\eta + \psi_\eta\psi'_\eta) = I_{n_y}.$$

We then find the expression,

$$\begin{aligned} & \sum_{\nu=0}^{\infty} tr\{\gamma(\nu+1)\theta\theta'_\nu\} \\ = & \sum_{\nu=0}^{\infty} tr\{\Sigma^{\frac{1}{2}}P'^{\nu+1}\Sigma^{-\frac{1}{2}}\Sigma^{-\frac{1}{2}}(I_{n_y}-P)^{-1}B\alpha(\alpha'\Omega^{-1}\alpha)^{-1}\alpha'B'P'^{\nu}\Sigma^{-\frac{1}{2}}\} \\ = & \ tr\{\Sigma^{-1}(I_{n_y}-P)^{-1}B\alpha(\alpha'\Omega^{-1}\alpha)^{-1}\alpha'B'(I_{n_y}-P')^{-1}(I_{n_y}+P')^{-1}P'\} \\ = & \ tr\{V'(I_{n_y}+P')^{-1}P'\} = tr\{P(I_{n_y}+P)^{-1}V\}, \end{aligned}$$

since

$$(I_{n_y}-P)^{-1}B\alpha = -(I_{n_u}, 0, \dots, 0)',$$

where we have used (50). A different expression can be found if the matrix P can be diagonalized. If we let $\Lambda = diag(\rho_1, \ldots, \rho_{n_y})$ and $K = (v_1, \ldots, v_{n_y})$ denote eigenvalues and eigenvectors of the matrix P, then $P = K\Lambda K^{-1}$. Note that these can be found from the eigenvalues for which $\rho \neq 1$ and the corresponding eigenvectors of the companion matrix.

In terms of these we have

$$\sum_{\nu=0}^{\infty} tr\{\gamma(\nu+1)\theta\theta'_\nu\} = tr\{P(I_{n_y}+P)^{-1}V\} = \sum_{j=1}^{n_y} \frac{\rho_j}{1+\rho_j}(K^{-1}VK)_{jj}.$$

$$\begin{aligned}
& \sum_{\nu=0}^{\infty} tr\{\gamma(\nu+1)\}tr\{\theta\theta'_\nu\} \\
= & \sum_{\nu=0}^{\infty} tr\{P'^{\nu+1}\}tr\{P'^{\nu}\Sigma^{-1}(I_{n_y}-P)^{-1}B\alpha(\alpha'\Omega^{-1}\alpha)^{-1}\alpha'B'\} \\
= & \sum_{\nu=0}^{\infty} tr\{P'^{\nu+1}\}tr\{P'^{\nu}V'(I_{n_y}-P')\} \\
= & \sum_{\nu=0}^{\infty} tr\{(P'\otimes P')^{\nu}(P'\otimes V'(I_{n_y}-P')\} \\
= & \; tr\{(I_{n_y}\otimes I_{n_y}-(P'\otimes P'))^{-1}(P'\otimes V'(I_{n_y}-P')\} \\
= & \; tr\{(P\otimes(I_{n_y}-P)V)(I_{n_y}\otimes I_{n_y}-(P\otimes P))^{-1}\}.
\end{aligned}$$

In terms of the eigenvectors and eigenvalues we can write

$$\begin{aligned}
\sum_{\nu=0}^{\infty} tr\{\gamma(\nu+1)\}tr\{\theta\theta'_\nu\} &= \sum_{\nu=0}^{\infty} tr\{P^{\nu+1}\}tr\{P^{\nu}(I_{n_y}-P)V\} \\
&= \sum_{\nu=0}^{\infty}\sum_{j,i=1}^{n_y} \rho_j^{\nu}\rho_i^{\nu}(1-\rho_i)(K^{-1}VK)_{ii} \\
&= \sum_{j,i=1}^{n_y} \frac{1-\rho_i}{1-\rho_i\rho_j}(K^{-1}VK)_{ii}
\end{aligned}$$

By the same method we find

$$\sum_{\nu=0}^{\infty} tr\{M^{\nu+1}\}tr\{\theta\theta'_\nu\} = tr\{(M\otimes(I_{n_y}-P)V)(I_{n_d}\otimes I_{n_y}-(M\otimes P))^{-1}\}.$$

3.3 The main result on a simple hypothesis

We now formulate the main result for test for a simple hypothesis on ξ_1 in the equation (15)

Theorem 3 *If the underlying process X_t is an AR(k) model with deterministic term Φd_t, the correction factor to the likelihood ratio test statistic for a simple hypothesis on the cointegrating space in (2.13) is given by*

$$\frac{E[-2\log LR|\alpha'_\perp\varepsilon]}{n_u n_a} \frac{1}{=} 1 + \tfrac{1}{T}[(n_d+n_y+n_b) + \tfrac{1}{2}(n_u+n_a+1)] + \tfrac{1}{Tn_u}[(n_a-1)v + 2(c+c_d)].$$

Here the constants v, c, and c_d are given by

$$v = tr\{V\} \quad (52)$$

$$c_d = tr\{(M \otimes (I_{n_y} - P)V)(I_{n_d} \otimes I_{n_y} - (M \otimes P))^{-1}\} \quad (53)$$

$$c = tr\{P(I_{n_y} + P)^{-1}V\} + tr\{[P \otimes (I_{n_y} - P)V][(I_{n_y} \otimes I_{n_y} - (P \otimes P))^{-1}]\} \quad (54)$$

Theorem 4 *where the matrix V is given by (51), P by (46), and M is defined by $d_{t+1} = Md_t$.*

4 Results for the autoregressive model

This section contains the consequences for the cointegrated vector autoregressive model for the various tests for linear hypotheses. We consider the autoregressive model with dimension n, lag length k, cointegrating rank r. In the following we restrict the deterministic term to be powers, like $d_t = 1$, $d_t = (1, t)$ or even $d_t = (1, t, t^2)$, since in this case one can see that $tr\{M^\nu\} = n_d$, the number of deterministic terms. We find, see (53), $c_d = n_d \sum_{\nu=0}^{\infty} tr\{\theta_\nu \theta'\} = n_d v$. The results can be modified if for instance a seasonal dummy is included, in which case $tr\{M^\nu\}$ is not constant but oscillates.

4.1 Test for simple hypothesis on the cointegrating space

The simple hypothesis on the cointegrating space is formulated as $\beta = H\phi$, where $H(n \times r)$ is known and $\phi(r \times r)$ is unknown.

Corollary 5 *In the $AR(k)$ model with power deterministic term Φd_t the correction term to the likelihood ratio test statistic for a simple hypothesis on the cointegrating space can be expressed in terms of the constants c and v, see (54) and (52):*

$$\frac{E[-2\log LR|\alpha_\perp'\varepsilon]}{r(n-r)} \overset{1}{=} 1 + \frac{1}{T}[(n_d + kn) + \tfrac{1}{2}(n+1)] + \frac{1}{Tr}[(n - r + 2n_d - 1)v + 2c].$$

Proof. This follows from Theorem 3 with the values $n_u = r$, $n_b = n_a = n - r$, $n_y = r + (k-1)n$, and using the result that $c_d = n_d v$. ∎

As a particularly simple case which is convenient for the simulation we consider

Corollary 6 *For the autoregressive model in n dimensions with 1 lag and $r = 1$ and constant term we find the correction factor*

$$\frac{E[-2\log LR|\alpha_\perp'\varepsilon]}{(n-1)} \overset{1}{=} 1 + \frac{1}{T}[\frac{3(n+1)}{2} - \frac{\alpha'\beta[(2+\alpha'\beta)n + 4(1+\alpha'\beta)]}{\beta'\Omega\beta\alpha'\Omega^{-1}\alpha}].$$

If the constant is absent we have

$$\frac{E[-2\log LR|\alpha_\perp'\varepsilon]}{(n-1)} \overset{1}{=} 1 + \frac{1}{T}[\frac{3n+1}{2} - \frac{\alpha'\beta[(2+\alpha'\beta)(n-2) + 4(1+\alpha'\beta)]}{\beta'\Omega\beta\alpha'\Omega^{-1}\alpha}].$$

Proof. In the special case of an autoregressive model with 1 lag and $r = 1$ we find that $P = 1 + \beta'\alpha$ and $M = 1$ or 0, and that

$$v = -\frac{\alpha'\beta(2+\alpha'\beta)}{\beta'\Omega\beta\alpha'\Omega^{-1}\alpha}, \; c = -2\frac{\alpha'\beta(1+\alpha'\beta)}{\beta'\Omega\beta\alpha'\Omega^{-1}\alpha}. \quad (55)$$

The result then follows from Corollary 5. ∎

4.2 Test for linear restrictions on β

We consider here hypotheses of the form $\beta = H\phi$, where $H(n \times s)$ is known and $\phi(s \times r)$ is unknown.

Calculation of the expectation of the log-likelihood ratio statistic is simply performed by applying Proposition 2 twice. The expectation of the second term is given by Corollary 5 and the expectation of the first can be found by replacing $(n-r)$ by $(s-r)$ as argued in subsection 2.3. We then find

Corollary 7 *The correction factor for the hypothesis $\beta = H\phi$ in the vector autoregressive model with power determinstic terms is given by*

$$\frac{E[-2\log LR|\alpha'_{\perp}\varepsilon]}{r(n-s)}$$
$$\overset{1}{=} 1 + \tfrac{1}{T}[(n_d + kn) + \tfrac{1}{2}(n+1+s-r)] + \tfrac{1}{Tr}[(n-2r+s+2n_d-1)v + 2c]$$

5 A simulation experiment

We simulate the distribution of the test statistic for an $n-$dimensional autoregressive model with one lag and one cointegrating relation and constant term in order to investigate if the correction factor improves the finite sample approximation. When simulating the distribution of the test statistics in the model

$$\Delta X_t = \alpha\beta' X_{t-1} + \mu + \varepsilon_t, \tag{56}$$

we have to specify the value of the parameters $(\alpha, \beta, \mu, \Omega)$, and the initial values, giving a total of $n + (n-1) + n + \frac{1}{2}n(n+1)$ parameters. We can, however, use the invariance of the statistic to reduce the number of parameters necessary to specify.

For all $n \times n$ matrices L of full rank, the transformation $Y = LX$ leaves the statistic invariant and corresponds to a change of the parameters into $(L\alpha,\ L^{-1\prime}\beta, L\Omega L', L\mu)$. We can reduce the number of parameters in the simulation experiment, by first choosing $L = \Omega^{-\frac{1}{2}}$ such that $\tilde{\varepsilon}_t = \Omega^{\frac{1}{2}}\varepsilon_t$ are i.i.d. $N_n(0, I_n)$. We can still transform by orthogonal transformations without changing the independence of the errors, and since there is no loss of generality in assuming that $\beta'\beta = 1$, we can rotate the coordinate system such that $\beta = (1, 0, \ldots, 0)' \in R^n$ and finally we rotate the coordinates $(2, \ldots, n)$ such that $\alpha = (\eta, \xi e_1')'$ where $e_1 = (1, 0, \ldots, 0)' \in R^{n-1}$. Finally we note that the values ξ and $-\xi$ give the same value of the test statistics and the correction. We therefore choose both η and ξ non-positive. Finally by a further rotation we can reduce the number of parameters in μ to three. Thus only five parameters need to be chosen in the simulation experiment when generating the data under the null hypothesis.

A more formal way of making this transformation in model (56) is to choose vectors $v_1 = \beta(\beta'\Omega\beta)^{-\frac{1}{2}}$,

$$v_2 = -(\Omega^{-1} - \beta(\beta'\Omega\beta)^{-1}\beta')\alpha\left(\alpha'\Omega^{-1}\alpha - \alpha'\beta(\beta'\Omega\beta)^{-1}\beta'\alpha\right)^{-\frac{1}{2}},$$

and $v_3, \ldots, v_n$ such that

$$v_i'\Omega v_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}.$$

For the new variables $Y_t = v'X_t$, the constant term is $\tilde{\mu} = v'\mu$, and the errors are $\tilde{\varepsilon}_t = v'\varepsilon_t$, which are i.i.d. $N_n(0, I_n)$. We get the equations

$$\begin{aligned}\Delta Y_{1t} &= \eta Y_{1t-1} + \tilde{\mu}_1 + \tilde{\varepsilon}_{1t}\\ \Delta Y_{2t} &= \xi Y_{1t-1} + \tilde{\mu}_2 + \tilde{\varepsilon}_{2t}\\ \Delta Y_{it} &= \tilde{\mu}_i + \tilde{\varepsilon}_{it}, i = 3, \dots, n,\end{aligned}$$

where

$$\begin{aligned}\eta &= (\beta'\Omega\beta)^{-\frac{1}{2}}\beta'\alpha(\beta'\Omega\beta)^{\frac{1}{2}}\\ \xi &= -\left(\alpha'\Omega^{-1}\alpha - \alpha'\beta(\beta'\Omega\beta)^{-1}\beta'\alpha\right)^{\frac{1}{2}}(\beta'\Omega\beta)^{\frac{1}{2}}.\end{aligned}$$

In the simulations we construct the stationary process $\beta'X_t$ by starting it at 0, and discarding the first 200 observations. The random walk $\alpha'_\perp X_t$ is started with initial value $\alpha'_\perp X_0 = 0$. Finally the process is constructed as

$$X_t = \alpha(\beta'\alpha)^{-1}\beta'X_t + \beta_\perp(\alpha'_\perp\beta_\perp)^{-1}\alpha'_\perp X_t,\ t = 1, \dots, T.$$

We thus want to test $\tau_2 = \cdots = \tau_n = 0$ in the model

$$\begin{aligned}\Delta X_{1t} &= \eta(X_{1t-1} + \tau_2 X_{2t-1} + \cdots + \tau_n X_{nt-1}) + \mu_1 + \varepsilon_{1t},\\ \Delta X_{2t} &= \xi(X_{1t-1} + \tau_2 X_{2t-1} + \cdots + \tau_n X_{nt-1}) + \mu_2 + \varepsilon_{2t},\\ \Delta X_{it} &= \mu_i + \varepsilon_{it}, i = 3, \dots, n,\end{aligned}$$

where ε_t are i.i.d. $N_n(0, I_n)$. The test is asymptotically distributed as $\chi^2(n-1)$, and under the null hypothesis we have the parameters

$$\alpha = \begin{pmatrix}\eta\\ \xi e_1\end{pmatrix}, \beta = \begin{pmatrix}1\\ 0\end{pmatrix}, \mu, \Omega = I_n.$$

Under the null hypothesis the system generates $I(1)$ variables with one cointegrating vector if $-2 < \eta < 0$. If $\eta = 0, \xi = 0$ then we get an $I(1)$ process with no cointegration and finally if $\eta = 0$ and $\xi \neq 0$ then the system will generate an $I(2)$ process.

The correction factor is given in (7) with $r = 1, k = 1, n_d = 1$, and

$$v = -\frac{\eta(2+\eta)}{\xi^2+\eta^2},\ c = -2\frac{\eta(1+\eta)}{\xi^2+\eta^2},$$

see (55), and is given by

$$BF = 1 + \frac{1}{T}\left(\frac{3(n+1)}{2} - \frac{\eta[(2+\eta)n + 4(1+\eta)]}{\eta^2+\xi^2}\right).$$

Note the special case of $\xi = 0$, where the correction factor has a pole at $\eta = 0$ corresponding to an extra unit root in the process. At this boundary the limit distribution is no longer χ^2 and a different type of correction could be calculated under the local alternative that $\eta \to 0$, $T \to \infty$, and $T\eta$ is fixed. Note also that if $\xi \neq 0$, and $\eta = 0$ then there is no singularity in the expression even though the process is $I(2)$.

$\xi \backslash \eta$	-0.1	-0.2	-0.4	-0.6	-0.8	-1.0
0.0	$\frac{0.0}{35.0}$	$\frac{1.1}{29.4}$	$\frac{3.1}{20.7}$	$\frac{4.8}{16.5}$	$\frac{5.3}{14.5}$	$\frac{5.4}{12.2}$
-0.1	$\frac{0.8}{30.6}$	$\frac{1.8}{27.0}$	$\frac{3.5}{19.6}$	$\frac{4.5}{15.9}$	$\frac{5.3}{13.9}$	$\frac{5.1}{12.4}$
-0.2	$\frac{4.0}{23.1}$	$\frac{3.3}{22.3}$	$\frac{4.0}{18.9}$	$\frac{4.7}{16.1}$	$\frac{5.1}{13.5}$	$\frac{5.1}{11.6}$
-0.4	$\frac{6.5}{15.8}$	$\frac{5.7}{17.1}$	$\frac{4.8}{16.4}$	$\frac{5.2}{15.0}$	$\frac{5.2}{13.1}$	$\frac{5.8}{12.1}$
-0.6	$\frac{6.5}{13.0}$	$\frac{5.8}{14.1}$	$\frac{5.5}{14.3}$	$\frac{5.8}{13.9}$	$\frac{5.8}{12.2}$	$\frac{5.7}{11.9}$
-0.8	$\frac{6.5}{12.2}$	$\frac{6.2}{12.3}$	$\frac{5.6}{12.6}$	$\frac{5.7}{13.0}$	$\frac{5.8}{12.7}$	$\frac{5.4}{11.3}$
-1.0	$\frac{6.1}{11.4}$	$\frac{5.7}{11.4}$	$\frac{5.7}{12.2}$	$\frac{5.8}{12.1}$	$\frac{5.5}{11.5}$	$\frac{5.5}{11.1}$

Table 1: Simulation of $T = 50$ observations from an $AR(1)$ process in 5 dimensions with $r = 1$ cointegrating relation. Number of simulations is 10.000. The table gives the corrected p-value over the uncorrected p-value for a nominal 5% test. The constant term is $\mu = 5$. The simulation standard error is 0.2%.

For the simulations we take $T = 50$ observations in an $n = 5$ dimensional system, and simulate the distribution of the test statistic 10.000 times. The simulations are performed in RATS. The parameter η ranges from -0.1 to -1.0, and the parameter ξ from 0.0 to -1.0. The results do not depend much on the value of μ so we take $\mu_i = 5.0$.

For each value of ξ and η the entries in Table 5.1 are

$$\frac{\text{corrected } p-\text{value of a nominal 5\% test}}{\text{actual } p-\text{value of a nominal 5\% test}},$$

that is, we take the 95% quantile from the $\chi^2(4)$ distribution and calculate the actual $p-$value as the frequency of rejections. We then correct the test statistic by the correction factor and calculate the rejection frequency again. This gives the corrected $p-$value. It is seen from Table 5.1, that the actual size can be quite distorted. For small values of η and ξ we get a size of up to 35%. The correction brings that down to 0.0% which is clearly too much, but for the combinations of η and ξ corresponding to $\xi^2 + \eta^2 < 0.17$, the correction gives a size of about 4-5%. Thus for these values the correction works quite well.

Note that when $\xi \neq 0$, the asymptotics seems to work even for small η. This is consistent with the fact that for $\eta \to 0$, $\xi \neq 0$ the model converges to a model for an $I(2)$ variable, but we know that the usual test on β remain valid even under $I(2)$ assumptions, see Johansen (1995) and Paruolo (1996, 1999), even though the interpretation changes.

If $\xi + \eta > -0.4$ the results are not so good, but notice that when $\eta = -0.10$, and $T = 50$ then $T\eta = -5$. The asymptotic power of the test for cointegrating rank is very low in such a case, and in most such situations the question of hypothesis testing on β would not be relevant. For $T = 100$ the results are given in Table 2, and it is seen that the parameter values for which the approximation is not so good have been reduced to $\xi^2 + \eta^2 < 0.05$.

$\xi\backslash\eta$	-0.1	-0.2	-0.4	-0.6	-0.8	-1.0
0.0	$\frac{0.5}{24.4}$	$\frac{2.5}{17.8}$	$\frac{4.2}{12.3}$	$\frac{4.8}{10.4}$	$\frac{5.0}{8.5}$	$\frac{5.0}{8.3}$
-0.1	$\frac{2.4}{19.6}$	$\frac{3.3}{16.6}$	$\frac{4.3}{12.0}$	$\frac{5.1}{10.2}$	$\frac{5.2}{8.9}$	$\frac{5.2}{8.3}$
-0.2	$\frac{4.7}{13.8}$	$\frac{4.2}{13.1}$	$\frac{4.8}{12.0}$	$\frac{5.0}{9.9}$	$\frac{5.3}{9.1}$	$\frac{5.1}{8.4}$
-0.4	$\frac{5.6}{9.4}$	$\frac{4.9}{10.0}$	$\frac{5.3}{10.2}$	$\frac{4.9}{9.1}$	$\frac{5.1}{8.7}$	$\frac{5.1}{8.1}$
-0.6	$\frac{5.1}{7.6}$	$\frac{5.0}{8.4}$	$\frac{4.9}{9.2}$	$\frac{4.8}{8.4}$	$\frac{5.3}{8.2}$	$\frac{5.3}{7.7}$
-0.8	$\frac{5.1}{7.7}$	$\frac{5.4}{8.4}$	$\frac{5.2}{8.4}$	$\frac{5.0}{8.3}$	$\frac{5.4}{8.6}$	$\frac{5.2}{7.7}$
-1.0	$\frac{5.2}{7.3}$	$\frac{5.2}{7.4}$	$\frac{5.1}{8.3}$	$\frac{5.3}{8.3}$	$\frac{5.0}{7.6}$	$\frac{5.2}{7.8}$

Table 2: Simulation of $T = 100$ observations from an $AR(1)$ process in 5 dimensions with $r = 1$ cointegrating relation. Number of simulations is 10.000. The table gives the corrected p-value over the uncorrected p-value for a nominal 5% test. The constant term is $\mu = 5$. The simulation standard error is 0.2%.

6 Conclusion

This paper has demonstrated that it is possible in vector autoregressive model with known adjustment coefficients to derive a correction term to the likelihood ratio test for some linear hypotheses on β. The correction factor depends on the sample size and the parameters. It is seen that if the adjustment coefficients are small the usual χ^2 approximation needs an improvement. By a few simulations it is indicated that when the adjustment coefficients have a reasonable size the approximation to the asymptotic distribution is improved. The model with unknown α is investigated in another paper, see Johansen (1999).

7 Acknowledgments

I would like to thank Bent Nielsen for insightful comments and Thomas Rothenberg for the reference to the work of Cavanagh.

8 References

Ahn, S. K. and C. G. Reinsel (1990), Estimation for partially non-stationary multivariate autoregressive models, *Journal of the American Statistical Association* 85, 813-823.

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

Cavanagh, C. (1983), Hypothesis testing in models with discrete dependent variables, Ph.D. Thesis, University of California, Berkeley.

Cribari-Neto, F. and Cordeiro, G. (1996), On Bartlett and Bartlett-type corrections, *Econometric Reviews* 15, 339-367.

Fachin, S. (1997) Bootstrap and asymptotic tests of long-run relationships in cointegrated systems, Discussion paper, University of Rome.

Gredenhoff, M. and Jacobson, T. (1998), Bootstrap testing and approximate finite sample distributions for test of linear restrictions on cointegrating vectors, Discussion paper, Stockholm School of Economics.

Jacobson, T. (1995), Simulating small sample properties of the maximum likelihood cointegration model: estimation and testing, *Finnish Economic Papers* 8, 96-107.

Jacobson, T. Vredin, A., and Warne, A. (1998), Are real wages and unemployment related ?, *Economica* Forthcoming.

Johansen, S. (1988), Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12, 231-254.

A Statistical Analysis of Cointegration for $I(2)$ Variables. *Econometric Theory* 11, 25–59 (1995)

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

Johansen, S. (1999), A Bartlett correction factor for tests on the cointegrating relations. Discussion paper.

Kiviet, J. F. and Phillips, G. D. A. (1997a), Bias and mean squared error of least squares estimators in dynamic regression model with a unit root. Discussion paper, Tinbergen Institute.

Kiviet, J. F. and Phillips, G. D. A. (1997b), Higher-order asymptotic expansion of the least-squares estimation bias in first-order dynamic regression models, Discussion paper, Tinbergen Institute.

Paruolo, P. (1999): Asymptotic efficiency of the 2 step estimator in $I(2)$ VAR systems. To appear in *Econometric Theory.*

Paruolo, P. (1996): On the determination of integration indices in $I(2)$ systems. *Journal of Econometrics,* 72, 313-356.

Phillips, P.C.B. (1991), Optimal inference in cointegrated systems, *Econometrica* 59, 283-306.

Rothenberg, T. J. (1988), Approximate power functions for some robust tests of regression coefficients, *Econometrica, 1*56*,* 997-1019.

9 Appendix: Proof of Proposition (1)

This section contains the proof of Proposition 1, and for ease of reference we repeat some of the definitions here. From the process U_t, Y_t, a_{t-1}, and b_t one defines product moments by

$$\begin{aligned} N &= S_{ua} \\ B &= S_{ay}(T^{-1}S_{yy.b,1})^{-1}S_{ya}, \\ A &= (T^{-\frac{1}{2}}S_{uy.b,1})(T^{-1}S_{yy.b,1})^{-1}S_{ya}, \end{aligned}$$

The expansion of the likelihood ratio test statistic is then

$$\begin{aligned} & tr\{Q\} \\ = \; & tr\{(I_{n_u} + T^{-\frac{1}{2}}D_1 - T^{-1}D_2)^{-1}(N - T^{-\frac{1}{2}}A)'(I_{n_a} - T^{-1}B)^{-1}(N - T^{-\frac{1}{2}}A)\} \\ \stackrel{1}{=} \; & tr\{N'N - T^{-\frac{1}{2}}(N'A + A'N + D_1N'N) \\ + \; & T^{-1}(D_2N'N + D_1^2N'N + D_1(N'A + A'N) + N'BN + A'A)\}, \end{aligned}$$

We start by giving a result about moments in the multivariate Gaussian distribution.

Lemma 8 *Let U_i, $i = 1, \dots, T$, be i.i.d. $N_{n_u}(0, I_{n_u})$. For any $n_u \times n_u$ matrix M*

$$tr\{E[U_i'MU_j]\} = \begin{cases} tr\{M\} & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

$$tr\{E[U_iU_j'MU_kU_l']\} = \begin{cases} (n_u + 2)tr\{M\} & \text{if } l = i = j = k, \\ n_u tr\{M\} & \text{if } l = i \neq j = k, \\ tr\{M\}. & \text{if } l = j \neq i = k, \\ tr\{M\} & \text{if } l = k \neq i = j, \\ 0 & \text{otherwise.} \end{cases}$$

It follows that

$$\sum_{i,j,k,l} tr\{E[U_ib_{ij}U_j'MU_kd_{kl}U_l']\} = \sum_{i,j}(n_ub_{ij}d_{ji} + b_{jj}d_{ii} + b_{ji}d_{ji})tr\{M\}$$

Proof. The first is a property of the multivariate Gaussian distribution and the next follows by summation. ■

9.1 Proof of (37):

We have to prove that

$$tr\{E[(N'N)^2|\alpha_\perp'\varepsilon]\} \stackrel{0}{=} n_un_a(n_u + n_a + 1).$$

This follows from Lemma 8:

$$\begin{aligned}
& tr\{E[(N'N)^2|\alpha_\perp'\varepsilon]\} \\
= \; & tr\{E[S_{ua}S_{au}S_{ua}S_{au}|\alpha_\perp'\varepsilon]\} \\
= \; & \sum_{i,j,k,l} tr\{E[U_i a_{i-1}' a_{j-1} U_j' U_k a_{k-1}' a_{l-1} U_l']\} \\
= \; & \sum_{i,j} (n_u a_{i-1}' a_{j-1} a_{j-1}' a_{i-1} + a_{i-1}' a_{i-1} a_{j-1}' a_{j-1} + a_{i-1}' a_{j-1} a_{i-1}' a_{j-1}) tr\{I_{n_u}\} \\
= \; & n_u(n_u tr\{S_{aa}^2\} + tr^2\{S_{aa}\} + tr\{S_{aa}^2\}) \\
= \; & n_u(n_u n_a + n_a^2 + n_a) = n_u n_a (n_u + n_a + 1).
\end{aligned}$$

9.2 Proof of (38):

To prove

$$tr\{E[N'N|\alpha_\perp'\varepsilon]\} \overset{0}{=} n_u n_a$$

we just note that $tr\{N'N\}$ follows a χ^2 distribution with $n_u n_a$ degrees of freedom such that the result follows.

9.3 Proof of (39):

The main difficulty is to prove (39), that is

$$\begin{aligned}
& T^{\frac{1}{2}} tr\{E[AN'|\alpha_\perp'\varepsilon]\} \overset{0}{=} n_a(n_u+1) tr\{\theta\theta'\} + n_u tr\{\mathcal{I}_{ba}\mathcal{I}_{ab}\psi'\psi\} \\
& -n_a \sum_{\nu=0}^{\infty} [tr\{\gamma(\nu+1)\theta\theta_\nu'\} + (tr\{\gamma(\nu+1)\} + tr\{M^{\nu+1}\}) tr\{\theta\theta_\nu'\}].
\end{aligned}$$

We have to investigate the product

$$T^{\frac{1}{2}} tr\{E[AN'|\alpha_\perp'\varepsilon]\} = tr\{E[S_{uy.b,d}(T^{-1}S_{yy.b,d})^{-1}S_{ya}S_{au}|\alpha_\perp'\varepsilon]\}.$$

The first factor is $S_{uy.b,d} = S_{uy} - S_{ud}S_{dd}^{-1}S_{dy} - S_{ub.d}S_{bb.d}^{-1}S_{by.d}$. All terms may give a contribution. We also have

$$T^{-1}S_{yy.b,d} = T^{-1}S_{yy} - T^{-1}S_{yd}S_{dd}^{-1}S_{dy} - T^{-1}S_{yb.d}S_{bb.d}^{-1}S_{by.d},$$

which shows that we can replace $T^{-1}S_{yy.b,d}$ by $T^{-1}S_{yy}$, since the difference is of the order of T^{-1}.

We next show that the expectation $E[T^{-1}S_{yy}|\alpha_\perp'\varepsilon]$ converges to the identity by the normalizations we have imposed

$$
\begin{aligned}
E[T^{-1}S_{yy}|\alpha'_\perp\varepsilon] &= T^{-1}\sum_{t=1}^{T} E[Y_{t-1}Y'_{t-1}|\alpha'_\perp\varepsilon] \\
&= \sum_{i=0}^{\infty}\theta_i\theta_i' + \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\psi_i[T^{-1}\sum_{t=1}^{T} b_{t-1-i}b'_{t-1-j}]\psi_j' \\
&\to \sum_{i=0}^{\infty}\theta_i\theta_i' + \sum_{i=0}^{\infty}\psi_i\psi_i' = I_{n_y}. \qquad (57)
\end{aligned}
$$

We here used the assumption that the given quantities b_t satisfy (30), as well as (48).

Thus in the final result we can replace $E[T^{-1}S_{yy}|\alpha'_\perp\varepsilon]$ by the identity, and hence we shall leave it out in the following. We also apply the expansion

$$
\begin{aligned}
\left(T^{-1}S_{yy}\right)^{-1} &= \left(E[T^{-1}S_{yy}|\alpha'_\perp\varepsilon] + (T^{-1}S_{yy} - E[T^{-1}S_{yy}|\alpha'_\perp\varepsilon])\right)^{-1} \\
&= \left(E[T^{-1}S_{yy}|\alpha'_\perp\varepsilon]\right)^{-1} - (T^{-1}S_{yy} - E[T^{-1}S_{yy}|\alpha'_\perp\varepsilon]) + \cdots \\
&= I_{n_y} - (T^{-1}S_{yy} - E[T^{-1}S_{yy}|\alpha'_\perp\varepsilon]) + \cdots .
\end{aligned}
$$

We find that we have to evaluate the terms

$$
\begin{aligned}
& T^{\frac{1}{2}}tr\{E[AN'|\alpha'_\perp\varepsilon]\} \\
& \overset{0}{=} tr\{E[S_{uy}S_{ya}S_{au}|\alpha'_\perp\varepsilon]\} \qquad (58)\\
& -tr\{E[S_{ud}S_{dd}^{-1}S_{dy}S_{ya}S_{au}|\alpha'_\perp\varepsilon]\} \qquad (59)\\
& -tr\{E[S_{ub.d}S_{bb.d}^{-1}S_{by.d}S_{ya}S_{au}|\alpha'_\perp\varepsilon]\} \qquad (60)\\
& -tr\{E[S_{uy}(T^{-1}S_{yy} - E[T^{-1}S_{yy}|\alpha'_\perp\varepsilon])S_{ya}S_{au}|\alpha'_\perp\varepsilon]\}. \qquad (61)
\end{aligned}
$$

We take the terms one by one

(58): $tr\{E[S_{uy}S_{ya}S_{au}|\alpha'_\perp\varepsilon]\} \overset{0}{=} (n_u+1)n_a tr\{\theta'\theta\} + n_u tr\{\psi'\psi\mathcal{I}_{ba}\mathcal{I}_{ab}\}$

$$
\begin{aligned}
& tr\{E[S_{uy}S_{ya}S_{au}|\alpha'_\perp\varepsilon]\} \\
=& \sum_{i,j,k} tr\{E[U_iY'_{i-1}Y_{j-1}a'_{j-1}a_{k-1}U'_k|\alpha'_\perp\varepsilon]\} \\
=& \sum_{i,j,k}\sum_{\nu,\mu} tr\{E[U_i(\theta_v U_{i-1-v} + \psi_v b_{i-1-v})'(\theta_\mu U_{j-1-\mu} + \psi_\mu b_{j-1-\mu})a'_{j-1}a_{k-1}U'_k]\}.
\end{aligned}
$$

We get a contribution if the number of stochastic factors is either four or two.

For four stochastic factors we find

$$\sum_{i,j,k}\sum_{v,\mu} tr\{E[U_iU'_{i-1-\nu}\theta'_\nu\theta_\mu U_{j-1-\mu}a'_{j-1}a_{k-1}U'_k]\}.$$

For $i = k, i-1-v = j-1-\mu$ we get

$$
\begin{aligned}
& \sum_{i}\sum_{v,\mu} tr\{E[U_iU'_{i-1-\nu}\theta'_\nu\theta_\mu U_{i-1-v}a'_{i-1+\mu-v}a_{i-1}U'_i]\} \\
=& n_u\sum_{i}\sum_{v,\mu} tr\{\theta'_\nu\theta_\mu\}tr\{a'_{i-1+\mu-v}a_{i-1}\} \\
=& n_u n_a\sum_{v,\mu} tr\{\theta'_\nu\theta_\mu\} = n_u n_a tr\{\theta'\theta\}.
\end{aligned}
$$

For $i = j - 1 - \mu, i - 1 - v = k$ we find

$$\begin{aligned} & \sum_i \sum_{v,\mu} tr\{E[U_i U'_{i-1-\nu} \theta'_\nu \theta_\mu U_i a'_{i+\mu} a_{i-2-v} U'_{i-v-1}]\} \\ = \quad & n_a \sum_{v,\mu} tr\{\theta'_\mu \theta_v\} = n_a tr\{\theta' \theta\} \end{aligned}$$

Thus we find with four stochastic terms the contribution

$$n_a(n_u + 1) tr\{\theta' \theta\} \tag{62}$$

Next consider the term with just two stochastic terms

$$\sum_{i,j,k} \sum_{v,\mu} tr\{E[U_i b'_{i-1-\nu} \psi'_\nu \psi_\mu b_{j-1-\mu} a'_{j-1} a_{k-1} U'_k]\}.$$

We get a contribution only if $k = i$ and find

$$\begin{aligned} & n_u \sum_{i,j} \sum_{v,\mu} tr\{b'_{i-1-\nu} \psi'_\nu \psi_\mu b_{j-1-\mu} a'_{j-1} a_{i-1}\} \\ \to \quad & n_u tr\{\psi' \psi \mathcal{I}_{ba} \mathcal{I}_{ab}\}, \end{aligned} \tag{63}$$

where we have used (30). This proves (58).

(59): $-tr\{E[S_{ud} S_{dd}^{-1} S_{dy} S_{ya} S_{au} | \alpha'_\perp \varepsilon]\} \overset{0}{=} -n_a \sum_{\nu=0}^\infty tr\{M^{\nu+1}\} tr\{\theta_\nu \theta'\}$

$$\begin{aligned} & -tr\{E[S_{ud} S_{dd}^{-1} S_{dy} S_{ya} S_{au} | \alpha'_\perp \varepsilon]\} \\ = \quad & - \sum_{i,j,k,s} \sum_{\nu,\mu} tr\{E[U_i d'_i S_{dd}^{-1} d_s (U'_{s-1-v} \theta'_v + b'_{s-1-v} \psi'_v) \\ & \qquad \times (\theta_\mu U_{j-1-\mu} + \psi_\mu b_{j-1-\mu}) a'_{j-1} a_{k-1} U'_k]\}. \end{aligned}$$

Again we only have to consider terms with four or two stochastic factors. With four we find

$$\begin{aligned} & - \sum_{i,j,k,s} \sum_{\nu,\mu} tr\{E[U_i d'_i S_{dd}^{-1} d_s U'_{s-1-v} \theta'_v)(\theta_\mu U_{j-1-\mu}) a'_{j-1} a_{k-1} U'_k]\} \\ = \quad & - \sum_{i,j,k,s} \sum_{\nu,\mu} tr\{E[U_i d'_i S_{dd}^{-1} d_{s+1+v} U'_s \theta'_v \theta_\mu U_j a'_{j+\mu} a_{k-1} U'_k]\}. \end{aligned}$$

By Lemma 8 we find that this equals

$$\begin{aligned} & - \textstyle\sum_{\nu,\mu} \sum_{i,j} (n_u d'_i S_{dd}^{-1} d_{j+1+v} a'_{j+\mu} a_{i-1} + d'_i S_{dd}^{-1} d_{j+1+v} a'_{i+\mu} a_{j-1} \\ & \qquad + d'_i S_{dd}^{-1} d_{i+1+v} a'_{j+\mu} a_{j-1}) tr\{\theta'_\nu \theta_\mu\} \\ & = -n_a \textstyle\sum_\nu tr\{d'_i S_{dd}^{-1} M^{\nu+1} d_i\} tr\{\theta'_\nu \theta\} \\ & = -n_a \textstyle\sum_\nu tr\{M^{\nu+1}\} tr\{\theta'_\nu \theta\} \end{aligned} \tag{64}$$

since $\sum_i a_{i-1} d'_i = 0$. The term with two stochastic factors is

$$
\begin{aligned}
& - \sum_{i,j,k,s} \sum_{\nu,\mu} tr\{E[U_i d_i' S_{dd}^{-1} d_s b_{s-1-v}' \psi_v' \psi_\mu b_{j-1-\mu} a_{j-1}' a_{k-1} U_k']\} \\
= \quad & -n_u \sum_{i,j,s} \sum_{\nu,\mu} tr\{d_i' S_{dd}^{-1} d_s\} tr\{b_{s-1-v}' \psi_v' \psi_\mu b_{j-1-\mu} a_{j-1}' a_{i-1}\} \to 0,
\end{aligned}
$$

since $\sum_i a_{i-1} d_i' = 0$.

(60): $-tr\{E[S_{ub.d} S_{bb.d}^{-1} S_{by.d} S_{ya} S_{au} | \alpha_\perp' \varepsilon]\} \overset{0}{=} 0$

This term is of the order of a constant so we can replace $S_{ub.d} S_{bb.d}^{-1} S_{by.d}$ by $S_{ub} S_{bb}^{-1} S_{by}$, and $T^{-1} S_{bb}$ by I_{n_b} :

$$
\begin{aligned}
& -tr\{E[S_{ub.d} S_{bb.d}^{-1} S_{by.d} S_{ya} S_{au} | \alpha_\perp' \varepsilon]\} \overset{0}{=} -T^{-1} tr\{E[S_{ub} S_{by} S_{ya} S_{au} | \alpha_\perp' \varepsilon]\} \\
= \quad & -T^{-1} \sum_{i,j,k,s} \sum_{\nu,\mu} tr\{E[U_i b_i' b_s (U_{s-1-v}' \theta_v' + b_{s-1-v}' \psi_v')(\theta_\mu U_{j-1-\mu} + \psi_\mu b_{j-1-\mu}) a_{j-1}' a_{k-1} U_k']\}.
\end{aligned}
$$

Again we only have to consider terms with four or two stochastic factors. With four we find

$$
\begin{aligned}
& -\frac{1}{T} \sum_{s,i,j,k} \sum_{\nu,\mu} tr\{E[U_i b_i' b_s (U_{s-1-v}' \theta_v')(\theta_\mu U_{j-1-\mu}) a_{j-1}' a_{k-1} U_k']\}, \\
& -\frac{1}{T} \sum_{s,i,j,k} \sum_{\nu,\mu} tr\{E[U_i b_i' b_{s+1+v} U_s' \theta_v' \theta_\mu U_j a_{j+\mu}' a_{k-1} U_k']\},
\end{aligned}
$$

which by Lemma 8 equals

$$
\begin{aligned}
& -\frac{1}{T} \sum_{i,j} tr\{n_u b_i' b_{j+\nu+1} a_{j+\mu}' a_{i-1} + b_i' b_{i+1+\nu} a_{j+\mu}' a_{j-1} + b_i' b_{j+\nu+1} a_{i+\mu}' a_{j-1}\} tr\{\theta_v' \theta_\mu\} \\
\to \quad & 0,
\end{aligned}
$$

since $\mathcal{I}_{ab}$ is finite and $\frac{1}{T} \sum_i b_i' b_{i+1+\nu} \to 0$.

With two stochastic terms we find

$$
\begin{aligned}
= \quad & -\frac{1}{T} \sum_{i,j,k,s} \sum_{\nu,\mu} tr\{E[U_i b_i' b_s b_{s-1-v}' \psi_v' \psi_\mu b_{j-1-\mu} a_{j-1}' a_{k-1} U_k']\} \\
= \quad & -\frac{r}{T} \sum_{i,j,s} \sum_{\nu,\mu} tr\{b_i' b_s b_{s-1-v}' \psi_v' \psi_\mu b_{j-1-\mu} a_{j-1}' a_{i-1}\} \to 0,
\end{aligned}
$$

since $\sum_i a_{i-1} b_i' = 0$, since the regressors are orthogonal. Hence there is no contribution from (60).

(61) : $-tr\{E[S_{uy}(T^{-1} S_{yy} - E[T^{-1} S_{yy} | \alpha_\perp' \varepsilon]) S_{ya} S_{au} | \alpha_\perp' \varepsilon]\}$

$\overset{0}{=} -n_a [\sum_\nu tr\{\gamma(\nu+1) \theta \theta_\nu'\} + \sum_\nu tr\{\gamma(\nu+1)\} tr\{\theta \theta_\nu'\}]$

$$
\begin{aligned}
& -tr\{E[S_{uy}(T^{-1}S_{yy} - E[T^{-1}S_{yy}|\alpha'_\perp\varepsilon]))S_{ya}S_{au}|\alpha'_\perp\varepsilon]\} \\
= \quad & -T^{-1}tr\{E[S_{uy}(\sum_i(Y_{i-1}Y'_{i-1} - E[Y_{i-1}Y'_{i-1}|\alpha'_\perp\varepsilon]))S_{ya}S_{au}|\alpha'_\perp\varepsilon]\} \\
= \quad & -T^{-1}\sum_{i,j,k,l}\sum_{\nu,\mu,\xi,\eta} tr\{E[U_i(U'_{i-1-\eta}\theta'_\eta + b'_{i-1-\eta}\psi'_\eta) \\
& \times(\theta_v(U_{l-v}U'_{l-\mu} - \delta_{\nu\mu}I_r)\theta'_\mu + \theta_v U_{l-v}b'_{l-\mu}\psi'_\mu + \psi_\nu b_{l-\nu}U'_{l-\mu}\theta'_\mu) \\
& \times(\theta_\xi U_{j-1-\xi} + \psi_\xi b_{j-1-\xi})a'_{j-1}a_{k-1}U'_k]\}.
\end{aligned}
$$

There are three parentheses involving a stochastic term. Let (a, b, c) denote the number of stochastic components from each parenthesis. We have to investigate the combinations $(0, 1, 1), (1, 1, 0), (0, 2, 0)$, and $(1, 2, 1)$, with an even number of stochastic factors.

That gives the six contributions to the sum since the factor from the middle parenthesis can be chosen in two ways

$$
\begin{aligned}
&(0,1,1)^a \sum_{i,j,k,l}\sum_{\nu,\mu,\xi,\eta} U_i(b'_{i-1-\eta}\psi'_\eta)(\theta_v U_{l-\nu}b'_{l-\mu}\psi'_\mu)(\theta_\xi U_{j-1-\xi})a'_{j-1}a_{k-1}U'_k \\
&(0,1,1)^b \sum_{i,j,k,l}\sum_{\nu,\mu,\xi,\eta} U_i(b'_{i-1-\eta}\psi'_\eta)(\psi_\nu b_{l-\nu}U'_{l-\mu}\theta'_\mu)(\theta_\xi U_{j-1-\xi})a'_{j-1}a_{k-1}U'_k \\
&(1,2,1) \sum_{i,j,k,l}\sum_{\nu,\mu,\xi,\eta} U_i(U'_{i-1-\eta}\theta'_\eta)\theta_v(U_{l-\nu}U'_{l-\mu} - \delta_{\nu,\mu}I_r)\theta'_\mu(\theta_\xi U_{j-1-\xi})a'_{j-1}a_{k-1}U'_k \\
&(1,1,0)^a \sum_{i,j,k,l}\sum_{\nu,\mu,\xi,\eta} U_i(U'_{i-1-\eta}\theta'_\eta)(\psi_\nu b_{l-\nu}U'_{l-\mu}\theta'_\mu)(\psi_\xi b_{j-1-\xi})a'_{j-1}a_{k-1}U'_k \\
&(1,1,0)^b \sum_{i,j,k,l}\sum_{\nu,\mu,\xi,\eta} U_i(U'_{i-1-\eta}\theta'_\eta)(\theta_v U_{l-\nu}b'_{l-\mu}\psi'_\mu)(\psi_\xi b_{j-1-\xi})a'_{j-1}a_{k-1}U'_k \\
&(0,2,0) \sum_{i,j,k,l}\sum_{\nu,\mu,\xi,\eta} U_i(b'_{i-1-\eta}\psi'_\eta)\theta_v(U_{l-\nu}U'_{l-\mu} - \delta_{\nu,\mu}I_r)\theta'_\mu(\psi_\xi b_{j-1-\xi})a'_{j-1}a_{k-1}U'_k
\end{aligned}
$$

We take the terms one by one

$$
(0,1,1)^a : -T^{-1}\sum_{i,j,k,l}\sum_{\nu,\mu,\xi,\eta} tr\{E\{U_i(b'_{i-1-\eta}\psi'_\eta)(\theta_\nu U_{l-\nu}b'_{l-\mu}\psi'_\mu)(\theta_\xi U_{j-1-\xi})a'_{j-1}a_{k-1}U'_k]\}.
$$

Using the result in Lemma 8 one can show that the term equals

$$
\begin{aligned}
& -T^{-1}\sum_{\nu,\mu,\xi,\eta} tr\{\theta'_\nu\psi_\eta(\sum_i b_{i-1-\eta}b'_{i-\mu+\nu})\psi'_\mu\theta_\xi\}tr\{\sum_j a'_{j-1}a_{j-2-\xi}\} \\
& -T^{-1}\sum_{\nu,\mu,\xi,\eta} n_u tr\{\theta'_\nu\psi_\eta(\sum_i b_{i-1-\eta}a'_{i-1})(\sum_j a_{j-\nu+\xi}b'_{j-\mu})\psi'_\mu\theta_\xi\} \\
& -T^{-1}\sum_{\nu,\mu,\xi,\eta} tr\{\theta'_\xi\psi_\mu(\sum_j b_{j-\mu}a'_{j-1-\nu})(\sum_i a_{i+\xi}b'_{i-1-\eta})\psi'_\eta\theta_\eta\} \\
\to \quad & -n_a\sum_{\nu,\eta} tr\{\theta'_\nu\psi_\eta\psi'_{\nu+\eta+1}\theta\},
\end{aligned}
\tag{65}
$$

since $T^{-1}\sum_i b_{i-1-\eta}b'_{i-\mu+\nu} \to I_{n-r}$ if $\mu = \nu+\eta+1$ and $tr\{\sum_j a'_{j-1}a_{j-2-\xi}\} \to tr(S_{aa}) = n_a$. The last two terms gives no contribution since $\mathcal{I}_{ab}$ is finite.

$$(0,1,1)^b : -T^{-1}\sum_{i,j,k,l}\sum_{\nu,\mu,\xi,\eta} tr\{E[U_i(b'_{i-1-\eta}\psi'_\eta)(\psi_\nu b_{l-\nu}U'_{l-\mu}\theta'_\mu)(\theta_\xi U_{j-1-\xi})a'_{j-1}a_{k-1}U'_k]\}$$

A similar calculation shows that we get a contribution for $[i = l-\mu \neq j-1-\xi = k]$

$$\begin{aligned}
& -T^{-1}\sum_{i,j,k,l}\sum_{\nu,\mu,\xi,\eta} tr\{E[U_i b'_{i-1-\eta}\psi'_\eta\psi_\nu b_{i-\nu+\mu}U'_i\theta'_\mu\theta_\xi U_k a'_{k+\xi}a_{k-1}U'_k]\} \\
= \; & -T^{-1}\sum_{\nu,\mu,\eta} tr\{(\sum_i b_{i-\nu+\mu}b'_{i-1-\eta})\psi'_\eta\psi_\nu\}tr\{\theta'_\mu\theta\}tr\{\sum_k a'_{k+\xi}a_{k-1}\} \\
\to \; & -n_a\sum_{\mu,\eta} tr\{\psi'_\eta\psi_{\mu+\eta+1}\}tr\{\theta'_\mu\theta\}, \qquad (66)
\end{aligned}$$

since $T^{-1}\sum_i b_{i-\nu+\mu}b'_{i-1-\eta} \to I_{n-r}$ if $\nu = \mu+\eta+1$, and $tr\{\sum_k a'_{k+\xi}a_{k-1}\} \to tr\{S_{aa}\} = n_a$.

$$(1,2,1) : \sum_{i,j,k,l}\sum_{\nu,\mu,\xi,\eta} tr\{E[U_i(U'_{i-1-\eta}\theta'_\eta)\theta_v(U_{l-\nu}U'_{l-\mu} - \delta_{\nu,\mu}I_r)\theta'_\mu(\theta_\xi U_{j-1-\xi})a'_{j-1}a_{k-1}U'_k]\}$$

We only get a contribution for $[i = l-\nu,\ i-1-\eta = l-\mu,\ j-1-\xi = k]$ which implies that $[\mu = \nu+\eta+1]$ and gives

$$\begin{aligned}
& -T^{-1}\sum_{i,j}\sum_{\nu,\xi,\eta} tr\{E[U_iU'_{i-1-\eta}\theta'_\eta\theta_v U_iU'_{i-1-\eta}\theta'_{\nu+\eta+1}\theta_\xi U_j a'_{j-1}a_{j-2-\xi}U'_{j-1-\xi}]\} \\
= \; & -T^{-1}\sum_{i,j}\sum_{\nu,\xi,\eta} tr\{\theta'_\nu\theta_\eta\theta'_{\nu+\eta+1}\theta_\xi a'_{j-1}a_{j-2-\xi}\} \\
\to \; & -n_a tr\{\sum_{\nu,\eta}\theta'_\nu\theta_\eta\theta'_{\nu+\eta+1}\theta\}. \qquad (67)
\end{aligned}$$

Finally we also get for $[i = l-\mu, i-1-\eta = l-\nu, j-1-\xi = k]$ which implies that $[\nu = \mu+\eta+1]$ and the contribution

$$-n_a tr\{\sum_{\nu,\eta}\theta'_{\nu+\eta+1}\theta_\eta\}tr\{\theta'_\nu\theta\}. \qquad (68)$$

We next collect the contributions from (65), (66), (67), and (68) and find a contribution to the expectation

$$-n_a\sum_{\nu,\eta} tr\{(\psi_\eta\psi'_{\nu+\eta+1} + \theta_\eta\theta'_{\nu+\eta+1})\theta\theta'_\nu\} = -n_a\sum_\nu tr\{\gamma(\nu+1)\theta\theta'_\nu\}, \qquad (69)$$

$$-n_a\sum_{\eta,\nu} tr\{\psi_\eta\psi'_{\nu+\eta+1}+\theta_\eta\theta'_{\nu+\eta+1}\}tr\{\theta\theta'_\nu\} = -n_a\sum_{\nu} tr\{\gamma(\nu+1)\}tr\{\theta\theta'_\nu\}. \tag{70}$$

This completes the expression for $Ttr\{E[N'A|\alpha'_\perp\varepsilon]\}$, given by (62), (63), (64), (69), and (70) since the remaining terms give no contribution, as we shall now show:

$$(1,1,0)^a : T^{-1}\sum_{i,j,k,l}\sum_{\nu,\mu,\xi,\eta} tr\{E[U_i(U'_{i-1-\eta}\theta'_\eta)(\psi_\nu b_{l-\nu}U'_{l-\mu}\theta'_\mu)(\psi_\xi b_{j-1-\xi})a'_{j-1}a_{k-1}U'_k]\}$$

We find a contribution only for k and l linked to i, which leaves two summations. Both give something bounded since $\mathcal{I}_{ab}$ is finite, hence the limit is zero.

$$(1,1,0)^b : T^{-1}\sum_{i,j,k,l}\sum_{\nu,\mu,\xi,\eta} tr\{E[U_i(U'_{i-1-\eta}\theta'_\eta)(\theta_v U_{l-\nu}b'_{l-\mu}\psi'_\mu)(\psi_\xi b_{j-1-\xi})a'_{j-1}a_{k-1}U'_k]\}.$$

The same argument applies as for $(1,1,0)^a$

$$(0,2,0) : T^{-1}\sum_{i,j,k,l}\sum_{\nu,\mu,\xi,\eta} tr\{E[U_i b'_{i-1-\eta}\psi'_\eta\theta_v(U_{l-\nu}U'_{l-\mu}-\delta_{\nu,\mu}I_r)\theta'_\mu\psi_\xi b_{j+1+\xi}a'_{j-1}a_{k-1}U'_k]\}.$$

Only if $[i=k=l-\nu=l-\mu]$ can we get a contribution and then the sum over i and j are both finite. Thus the limit is zero.

9.4 Proof of (40):

Next we prove

$$tr\{E[A'A|\alpha'_\perp\varepsilon]\} \overset{0}{=} n_a n_u tr\{\theta\theta'\} + n_u tr\{\mathcal{I}_{ba}\mathcal{I}_{ab}\psi'\psi\}.$$

We have the expression

$$\begin{aligned} & tr\{E[A'A|\alpha'_\perp\varepsilon]\} \\ = \; & T^{-1}tr\{E[(S_{uy.b,d})(T^{-1}S_{yy.b,d})^{-1}S_{ya}S_{ay}(T^{-1}S_{yy.b,d})^{-1}(S_{yu.b,d})|\alpha'_\perp\varepsilon]\}. \end{aligned}$$

Because of the factor T^{-1} we can replace throughout the matrix $(S_{uy.b,d})(T^{-1}S_{yy.b,d})^{-1}$ by $(S_{uy.b})(T^{-1}S_{yy.b})^{-1}$ and find the expression

$$\begin{aligned} & tr\{E[A'A|\alpha'_\perp\varepsilon]\} \\ & \overset{0}{=} T^{-1}tr\{E[S_{uy}S_{ya}S_{ay}S_{yu}|\alpha'_\perp\varepsilon]\} \\ = \; & T^{-1}\sum_{i,j,k,l}\sum_{\mu,\xi,\eta,\nu} tr\{E[U_i(U'_{i-1-\mu}\theta'_\mu + b'_{i-1-\mu}\psi'_\mu)(\theta_\xi U_{j-1-\xi}+\psi_\xi b_{j-1-\xi}) \\ & a'_{j-1}a_{k-1}(U'_{k-1-\nu}\theta'_\nu + b'_{k-1-\nu}\psi'_\nu)(\theta_\eta U_{l-1-\eta}+\psi_\eta b_{l-1-\eta})U'_l]\}. \end{aligned} \tag{71}$$

There are here four parentheses and we can choose a stochastic or a deterministic term from each. We apply again the notation (a, b, c, d) to indicate the number of stochastic components in each sum below. We find that we shall investigate the combinations $(1,1,1,1)$, $(1,1,0,0)$, $(1,0,1,0)$, $(1,0,0,1)$, $(0,1,1,0)$, $(0,1,0,1)$, $(0,0,1,1)$, $(0,0,0,0)$ and find the terms

$$(1,1,1,1) \sum_{i,j,k,l} \sum_{\mu,\xi,\eta,\nu} tr\{E[U_i(U'_{i-1-\mu}\theta'_\mu)(\theta_\xi U_{j-1-\xi})a'_{j-1}a_{k-1}(U'_{k-1-\nu}\theta'_\nu)(\theta_\eta U_{l-1-\eta})U'_l]\}$$

$$(1,0,0,1) \sum_{i,j,k,l} \sum_{\mu,\xi,\eta,\nu} tr\{E[U_i(U'_{i-1-\mu}\theta'_\mu)(\psi_\xi b_{j-1-\xi})a'_{j-1}a_{k-1}(b'_{k-1-\nu}\psi'_\nu)(\theta_\eta U_{l-1-\eta})U'_l]\}$$

$$(0,1,1,0) \sum_{i,j,k,l} \sum_{\mu,\xi,\eta,\nu} tr\{E[U_i(b'_{i-1-\mu}\psi'_\mu)(\theta_\xi U_{j-1-\xi})a'_{j-1}a_{k-1}(U'_{k-1-\nu}\theta'_\nu)(\psi_\eta b_{l-1-\eta})U'_l]\}$$

$$(0,0,0,0) \sum_{i,j,k,l} \sum_{\mu,\xi,\eta,\nu} tr\{E[U_i(b'_{i-1-\mu}\psi'_\mu)(\psi_\xi b_{j-1-\xi})a'_{j-1}a_{k-1}(b'_{k-1-\nu}\psi'_\nu)(\psi_\eta b_{l-1-\eta})U'_l]\}$$

$$(1,1,0,0) \sum_{i,j,k,l} \sum_{\mu,\xi,\eta,\nu} tr\{E[U_i(U'_{i-1-\mu}\theta'_\mu)(\theta_\xi U_{j-1-\xi})a'_{j-1}a_{k-1}(b'_{k-1-\nu}\psi'_\nu)(\psi_\eta b_{l-1-\eta})U'_l]\}$$

$$(1,0,1,0) \sum_{i,j,k,l} \sum_{\mu,\xi,\eta,\nu} tr\{E[U_i(U'_{i-1-\mu}\theta'_\mu)(\psi_\xi b_{j-1-\xi})a'_{j-1}a_{k-1}(U'_{k-1-\nu}\theta'_\nu)(\psi_\eta b_{l-1-\eta})U'_l]\}$$

$$(0,1,0,1) \sum_{i,j,k,l} \sum_{\mu,\xi,\eta,\nu} tr\{E[U_i(b'_{i-1-\mu}\psi'_\mu)(\theta_\xi U_{j-1-\xi})a'_{j-1}a_{k-1}(b'_{k-1-\nu}\psi'_\nu)(\theta_\eta U_{l-1-\eta})U'_l]\}$$

$$(0,0,1,1) \sum_{i,j,k,l} \sum_{\mu,\xi,\eta,\nu} tr\{E[U_i(b'_{i-1-\mu}\psi'_\mu)(\psi_\xi b_{j-1-\xi})a'_{j-1}a_{k-1}(U'_{k-1-\nu}\theta'_\nu)(\theta_\eta U_{l-1-\eta})U'_l]\}$$

We take the terms one by one

$$(1,1,1,1) : T^{-1} \sum_{i,j,k,l} \sum_{\mu,\xi,\eta,\nu} tr\{E[U_i(U'_{i-1-\mu}\theta'_\mu)(\theta_\xi U_{j-1-\xi})a'_{j-1}a_{k-1}(U'_{k-1-\nu}\theta'_\nu)(\theta_\eta U_{l-1-\eta})U'_l]$$

For $[i = l, \mu = \eta, j-1-\xi = k-1-\nu]$ we find the contribution

$$n_u n_a tr\{(\sum_\nu \theta_\nu \theta'_\nu)\theta\theta'\}. \tag{72}$$

$$(1,0,0,1) : T^{-1} \sum_{i,j,k,l} \sum_{\mu,\xi,\eta,\nu} tr\{E[U_i(U'_{i-1-\mu}\theta'_\mu)(\psi_\xi b_{j-1-\xi})a'_{j-1}a_{k-1}(b'_{k-1-\nu}\psi'_\nu)(\theta_\eta U_{l-1-\eta})U'_l]$$

For $[l = i, \mu = \eta]$ we get the limit

$$n_u tr\{(\sum_\nu \theta_\nu \theta'_\nu)\psi \mathcal{I}_{ba}\mathcal{I}_{ab}\psi'\} \tag{73}$$

$$(0,1,1,0) : T^{-1} \sum_{i,j,k,l} \sum_{\mu,\xi,\eta,\nu} tr\{E[U_i(b'_{i-1-\mu}\psi'_\mu)(\theta_\xi U_{j-1-\xi})a'_{j-1}a_{k-1}(U'_{k-1-\nu}\theta'_\nu)(\psi_\eta b_{l-1-\eta})U'_l]$$

$[i=l, j-1-\xi = k-1-\nu]$ gives

$$n_u n_a tr\{(\sum_\nu \psi_\nu \psi_\nu')\theta\theta'\} \tag{74}$$

$$(0,0,0,0): T^{-1}\sum_{i,j,k,l}\sum_{\mu,\xi,\eta,\nu} tr\{E[U_i(b_{i-1-\mu}'\psi_\mu')(\psi_\xi b_{j-1-\xi})a_{j-1}'a_{k-1}(b_{k-1-\nu}'\psi_\nu')(\psi_\eta b_{l-1-\eta})U_l']]$$

$[i=l]$ gives

$$n_u tr\{\psi\mathcal{I}_{ba}\mathcal{I}_{ab}\psi'(\sum_\nu \psi_\nu\psi_\nu')\}. \tag{75}$$

Finally we note that from (72) and (74) we find a contribution to (71)

$$n_u n_a tr\{(\sum_\nu \theta_\nu\theta_\nu')\theta\theta'\} + n_u n_a tr\{(\sum_\nu \psi_\nu\psi_\nu')\theta\theta'\} = n_u n_a tr\{\theta\theta'\}, \tag{76}$$

since $\sum_\nu(\theta_\nu\theta_\nu' + \psi_\nu\psi_\nu') = I_{n_y}$, see (57).

Combining (73) and (75) we find that a contribution to (71) is

$$n_u tr\{\psi\mathcal{I}_{ba}\mathcal{I}_{ab}\psi'(\sum_\nu \psi_\nu\psi_\nu')\} + n_u tr\{(\sum_\nu \theta_\nu\theta_\nu')\psi\mathcal{I}_{ba}\mathcal{I}_{ab}\psi'\} = n_u tr\{\psi\mathcal{I}_{ba}\mathcal{I}_{ab}\psi'\}. \tag{77}$$

The expressions (76) and (77) give the final result for $tr\{E[A'A|\alpha_\perp'\varepsilon]\}$ since the remaining terms tend to zero, as we shall now see:

$$(1,1,0,0): T^{-1}\sum_{i,j,k,l}\sum_{\mu,\xi,\eta,\nu} tr\{E[U_i(U_{i-1-\mu}'\theta_\mu')(\theta_\xi U_{j-1-\xi})a_{j-1}'a_{k-1}(b_{k-1-\nu}'\psi_\nu')(\psi_\eta b_{l-1-\eta})U_l'].$$

For this to give a non-zero value the indices i,j,l are linked. The summation over i, and k are bounded because $\mathcal{I}_{ab}$ is finite, hence the limit is zero. Exactly the same argument can be applied for $(1,0,1,0)$, $(0,1,0,1)$, $(0,0,1,1)$.

9.5 Proof of (41):

We then prove

$$tr\{E[N'BN|\alpha_\perp'\varepsilon]\} \overset{0}{=} n_a(n_a+n_u+1)tr\{\theta\theta'\} + n_u tr\{\mathcal{I}_{ba}\mathcal{I}_{ab}\psi'\psi\}.$$

From the expression

$$tr\{E[N'BN|\alpha_\perp'\varepsilon]\} = tr\{E[S_{ua}S_{ay}(T^{-1}S_{yy.b,1})^{-1}S_{ya}S_{au}|\alpha_\perp'\varepsilon]\},$$

we see that we can replace $T^{-1}S_{yy.b,1}$ by I_{n_y} and find

$$\begin{aligned} & tr\{E[N'BN|\alpha'_\perp\varepsilon]\} \overset{0}{=} tr\{E[S_{ua}S_{ay}S_{ya}S_{au}|\alpha'_\perp\varepsilon]\} \\ = & \sum_{i,j,k,l}\sum_{v,\mu} tr\{E[U_i a'_{i-1}a_{j-1}(U'_{j-1-v}\theta'_v + b'_{j-1-v}\psi'_v)(\theta_\mu U_{k-1-\mu} + \psi_\mu b_{k-1-\mu})a'_{k-1}a_{l-1}U'_l]\} \end{aligned}$$

The contribution from the term with three U's will be zero and what remains are the two terms

$$\sum_{i,j,k,l}\sum_{v,\mu} tr\{E[U_i a'_{i-1}a_{j-1}U'_{j-1-v}\theta'_v\theta_\mu U_{k-1-\mu}a'_{k-1}a_{l-1}U'_l]\}$$
$$+\sum_{i,j,k,l}\sum_{v,\mu} tr\{E[U_i a'_{i-1}a_{j-1}b'_{j-1-v}\psi'_v\psi_\mu b_{k-1-\mu}a'_{k-1}a_{l-1}U'_l]\}.$$

The first is evaluated by Lemma 8 and we find

$$\begin{aligned} & \sum_{i,j,k,l}\sum_{v,\mu} tr\{E[U_i a'_{i-1}a_{j-1}U'_{j-1-v}\theta'_v\theta_\mu U_{k-1-\mu}a'_{k-1}a_{l-1}U'_l]\} \\ = & \sum_{i,j}\sum_{v,\mu} a'_{i-1}a_{j-1}a'_{j-1-\nu+\mu}a_{i-1}tr\{I_{n_u}\}tr\{\theta'_v\theta_k\} \\ & +\sum_{i,j}\sum_{v,\mu} a'_{i-1}a_{j-1}a'_{i+\mu}a_{j-2-\nu}tr\{\theta'_v\theta_k\} + \sum_{i,l}\sum_{v,\mu} a'_{i-1}a_{i+v}a'_{l+\mu}a_{l-1}tr\{\theta'_v\theta_k\} \\ \rightarrow & \; tr\{\theta'\theta\}(n_u tr\{S_{aa}S_{aa}\} + tr\{S_{aa}S_{aa}\} + tr\{S_{aa}\}tr\{S_{aa}\}) \\ = & \; tr\{\theta'\theta\}(n_u n_a + n_a + n_a^2) = n_a(n_u + n_a + 1)tr\{\theta'\theta\}. \end{aligned} \tag{78}$$

The term with two deterministic factors becomes

$$\begin{aligned} & \sum_{i,j,k,l}\sum_{v,\mu} tr\{E[U_i a'_{i-1}a_{j-1}b'_{j-1-v}\psi'_v\psi_\mu b_{k-1-\mu}a'_{k-1}a_{l-1}U'_l]\} \\ = & \; n_u\sum_{i,j,k}\sum_{v,\mu} tr\{a'_{i-1}a_{j-1}b'_{j-1-\nu}\psi'_v\psi_\mu b_{k-1-\mu}a'_{k-1}a_{i-1}\} \\ \rightarrow & \; n_u tr\{\mathcal{I}_{ba}\mathcal{I}_{ab}\psi'\psi\}. \end{aligned} \tag{79}$$

The result for $tr\{E[N'BN|\alpha'_\perp\varepsilon]\}$ is contained in (78) and (79).

9.6 Proof of (42):

To prove

$$T^{\frac{1}{2}}tr\{E[D_1N'N|\alpha'_\perp\varepsilon]\} \overset{0}{=} n_u n_a(n_u+1)$$

we note that

$$
\begin{aligned}
& T^{\frac{1}{2}} tr\{E[D_1 N'N|\alpha'_\perp \varepsilon]\} \\
= \; & tr\{E[(S_{uu} - TI_{n_u})S_{ua}S_{au}|\alpha'_\perp \varepsilon]\} \\
= \; & \sum_{j,k,l} tr\{E[(U_j U_j' - I_{n_u})U_k a'_{k-1} a_{l-1} U_l']\} \\
= \; & \sum_{j} tr\{E[(U_j U_j' - I_{n_u})U_j a'_{j-1} a_{j-1} U_j']\} \\
= \; & tr\{S_{aa}\} tr\{E[(U_j U_j' - I_{n_u})U_j U_j']\} \\
= \; & n_u n_a (n_u + 1).
\end{aligned}
$$

9.7 Proof of (45):

Next consider

$$
tr\{E[D_1^2 N'N|\alpha'_\perp \varepsilon]\} \overset{0}{=} n_u n_a (n_u + 1).
$$

We find

$$
\begin{aligned}
& tr\{E[D_1^2 N'N|\alpha'_\perp \varepsilon]\} \\
= \; & T^{-1} tr\{E[(S_{uu} - TI_{n_u})(S_{uu} - TI_{n_u})S_{ua}S_{au}|\alpha'_\perp \varepsilon]\} \\
= \; & T^{-1} \sum_{i,j,k,l} tr\{E[(U_i U_i' - I_{n_u})(U_j U_j' - I_{n_u})U_k a'_{k-1} a_{l-1} U_l']\}.
\end{aligned}
$$

Here we take $[k = l]$ since otherwise the a_{k-1} will sum to zero. We then have to take $[i = j]$ to get a non-zero contribution and find

$$
\begin{aligned}
& T^{-1} \sum_{i,k} tr\{E[(U_i U_i' - I_{n_u})(U_i U_i' - I_{n_u})U_k a'_{k-1} a_{k-1} U_k']\} \\
= \; & tr\{S_{aa}\} tr\{E[(U_i U_i' - I_{n_u})(U_i U_i' - I_{n_u})\} \\
= \; & n_u n_a (n_u + 1).
\end{aligned}
$$

9.8 Proof of (43):

To prove

$$
tr\{E[D_2 N'N|\alpha'_\perp \varepsilon]\} \overset{0}{=} n_u n_a (n_d + n_y + n_b)
$$

we evaluate as follows

$$
\begin{aligned}
& tr\{E[D_2 N'N|\alpha'_\perp \varepsilon]\} \\
& = tr\{E[(S_{ud}S_{dd}^{-1}S_{du} + S_{ub.d}S_{bb.d}^{-1}S_{bu.d} + S_{uy.b,d}S_{yy.b,d}^{-1}S_{yu.b,d})S_{ua}S_{au}|\alpha'_\perp \varepsilon]\}
\end{aligned}
$$

1.

$$
\begin{aligned}
& tr\{E[S_{ud}S_{dd}^{-1}S_{du}S_{ua}S_{au}|\alpha_\perp'\varepsilon]\} \\
= & \sum_{i,j,k,l} tr\{E[U_i d_i' S_{dd}^{-1} d_j U_j' U_k a_{k-1}' a_{l-1} U_l']\} \\
= & n_u n_a n_d,
\end{aligned}
$$

by Lemma 8.

2.

$$
\begin{aligned}
& tr\{E[S_{ub^0}S_{b^0b^0}^{-1}S_{b^0u}S_{ua}S_{au}]\} \\
& \overset{0}{=} T^{-1} tr\{E[S_{ub}S_{bu}S_{ua}S_{au}]\} \\
= & T^{-1} \sum_{i,j,k,l} tr\{E[U_i b_i' b_j U_j' U_k a_{k-1}' a_{l-1} U_l']\} \\
= & n_u n_a n_b,
\end{aligned}
$$

by Lemma 8.

3. In the last expression we replace $S_{uy.b,d}S_{yy.b,d}^{-1}S_{yu.b,d}$ by $S_{uy}S_{yy}^{-1}S_{yu} \overset{0}{=} T^{-1}S_{uy}S_{yu}$:

$$
\begin{aligned}
& T^{-1} tr\{E[S_{uy}S_{yu}S_{ua}S_{au}]\} \\
= & T^{-1} \sum_{i,j,s,t} tr\{E[U_i Y_{i-1}' Y_{j-1} U_j' U_s a_{s-1}' a_{t-1} U_t']\} \\
= & T^{-1} \sum_{i,j,s,t} \sum_{\nu,\mu} tr\{E[U_i (U_{i-1-\nu}'\theta_\nu')(\theta_\mu U_{j-1-\mu}) U_j' U_s a_{s-1}' a_{t-1} U_t']\} \\
& + T^{-1} \sum_{i,j,s,t} \sum_{\nu,\mu} tr\{E[U_i (b_{i-1-\nu}'\psi_\nu')(\psi_\mu b_{j-1-\mu}) U_j' U_s a_{s-1}' a_{t-1} U_t']\} \\
= & \sum_\nu n_u tr\{\theta_\nu'\theta_\nu\} tr\{S_{aa}\} + \sum_\nu n_u tr\{\psi_\nu'\psi_\nu\} tr\{S_{aa}\} \\
= & n_u tr\{S_{aa}\} \sum_\nu (tr\{\theta_\nu\theta_\nu'\} + tr\{\psi_\nu\psi_\nu'\}) = n_u tr\{S_{aa}\} tr\{I_{n_y}\} = n_u n_a n_y.
\end{aligned}
$$

Note that the lag length appears here in the form of the matrix I_{n_y}. Adding these contributions we find

$$
n_u n_a n_d + n_u n_a n_b + n_u n_a n_y = n_u n_a (n_d + n_b + n_y).
$$

9.9 Proof of (44):

We finally prove

$$
tr\{D_1 N' A|\alpha_\perp'\varepsilon]\}.
$$

We have

$$
tr\{D_1 N' A|\alpha_\perp'\varepsilon]\} = tr\{E[(T^{-1}S_{uu} - I_{n_u})S_{uy.b,1}(T^{-1}S_{yy.b.1})^{-1}S_{ya}S_{au}|\alpha_\perp'\varepsilon]\}
$$

We replace $S_{uy.b,1}$ by S_{uy} and $T^{-1}S_{yy.b.1}$ by I_{n_y} and find

$$\begin{aligned}
& tr\{E[(T^{-1}S_{uu} - I_{n_u})S_{uy}S_{ya}S_{au}|\alpha'_{\perp}\varepsilon]\} \\
= \; & T^{-1}tr\{E[(S_{uu} - TII_{n_u})S_{uy}S_{ya}S_{au}|\alpha'_{\perp}\varepsilon]\} \\
= \; & T^{-1}\sum_{i,j,k,l} tr\{E[(U_iU_i' - I_{n_u})U_jY_{j-1}'Y_{k-1}a_{k-1}'a_{l-1}U_l']\}.
\end{aligned}$$

We find the two contributions

$$\begin{aligned}
& T^{-1}\sum_{i,j,k,l}\sum_{\nu,\mu} tr\{E[(U_iU_i' - I_r)U_j(U_{j-1-\nu}'\theta_{\nu}')(\theta_{\mu}U_{k-1-\mu})a_{k-1}'a_{l-1}U_l']\} \\
& +T^{-1}\sum_{i,j,k,l}\sum_{\nu,\mu} tr\{E[(U_iU_i' - I_r)U_j(b_{j-1-\nu}'\psi_{\nu}')(\psi_{\mu}b_{k-1-\mu})a_{k-1}'a_{l-1}U_l']\}.
\end{aligned}$$

In the first term we must have the indices $[i,j,k,l]$ all tied to j, since $j \neq j-1-\nu$. This leaves one index to sum and the sum of the $a_{k-1}'a_{k-1}$ is bounded. Hence no contribution.

In the second term we must have $[l = j = i]$, which leaves a summation over j and k. Both of these are finite since $\mathcal{I}_{ab}$ is finite. Thus the limit is zero.

This completes the proof of Proposition 1.