THE ASYMPTOTIC DISTRIBUTION OF LIKELIHOOD RATIO TEST STATISTICS FOR COINTEGRATION IN UNSTABLE VECTOR AUTOREGRESSIVE PROCESSES

By Bent Nielsen[1]

Department of Economics, University of Oxford

Address for correspondence: Nuffield College, Oxford OX1 1NF, UK
Email: bent.nielsen@nuf.ox.ac.uk
Web: http://www.nuff.ox.ac.uk/users/nielsen

4 September 2000

When analysing cointegration in vector autoregressive models it is usually assumed that (i) the number of cointegrating relations is not smaller than what is tested for, (ii) the number of unit roots equals the number of common stochastic trends, and (iii) the remaining characteristic roots of the time series are stationary roots. Condition (iii) could be violated in data from hyper-inflationary economies and in connection with seasonally integrated data while condition (ii) is not satisfied by processes which are integrated of order two. Here it is proved that condition (iii) is redundant when determining the cointegration rank whereas both condition (ii) and (iii) are redundant when testing linear restrictions on the cointegrating vector.

1 Introduction

Cointegration analysis was originally suggested by Engle and Granger (1987) as a method for finding long-run equilibria among stochastically trending variables. This analysis is typically used in the context of difference stationary multivariate time series, where the levels of the series exhibit non-stationary behaviour but stationarity is achieved by differencing. In that setting it is often possible to find stationary linear combinations which can be associated as long-run economic equilibria. In a recent study by Juselius and Mladenovic (1999) of Yugoslavian hyper-inflation data it is found that cointegration analysis is a useful econometric tool despite the explosive behaviour of data and the consequent violation of the usual assumptions to cointegration analysis. In this paper Johansen's (1988, 1996) likelihood-based procedure for cointegration analysis is investigated with a view towards this issue. In particular it is found that the asymptotic results are valid for explosively growing variables.

[1]Discussions with Katarina Juselius and Zorica Mladenovic are gratefully acknowledged.

The available tests for cointegration hinge on the assumption that data are difference stationary. This can be formulated more precisely in terms of three conditions to the characteristic roots of the time series: (i) the number of cointegrating relations is not smaller than what is tested for, (ii) the number of unit roots equals the number of common stochastic trends, and (iii) the remaining characteristic roots are stationary roots. It is proved that the condition (iii) is redundant when determining the number of cointegration relations whereas both of the conditions (ii) and (iii) are redundant when testing linear restrictions on the cointegration parameters.

The findings for the multivariate cointegration model generalise a result for unit root tests by Nielsen (2000) although the proof is more involved. For the univariate case the intuition is as follows. The unit root hypothesis is tested in an autoregressive model of order $(k+1)$ by first regressing differences and lagged levels of the process on k lagged differences, and then finding the sample correlation of the residuals. In that case the conditions (i)-(iii) concern k parameters corresponding to the k lags which are eliminated by regression in the first step of the analysis. In the multivariate case a p-dimensional model of order $(k+1)$ is analysed. This is analysed as before by first regressing on k lagged differences and then finding the sample canonical correlations of the residuals. Now the conditions (i)-(iii) concern $kp+r$ characteristic roots, where r is the number of cointegrating relations. The first step of the statistical analysis only eliminates kp roots, whereas the remaining r roots relate to the cointegrating relations. In general the cointegrating relations are not necessarily stationary and this feature adds considerably to the complexity to the proof.

Section 2 presents the statistical model along with the statistical analysis and the results. Robustness with respect to martingale difference innovations is also discussed. The proofs follow in an Appendix and are based on ideas of Johansen (1988, 1996), Lai and Wei (1982, 1983, 1985) and Chan and Wei (1988).

The following notation is used: for a $(p \times r)$-matrix α where $p \geq r$ the $\{p \times (p-r)\}$-dimensional orthogonal complement is denoted $\alpha_\perp$, whereas $\overline{\alpha} = \alpha(\alpha'\alpha)^{-1}$ and $\alpha^{\otimes 2} = \alpha\alpha'$. For a symmetric square matrix α let $\lambda_{\min}(\alpha)$ and $\lambda_{\max}(\alpha)$ denote the largest and the smallest eigenvalues respectively. The norm $||\alpha||$ is the Euclidean norm, so $||\alpha||^2 = \lambda_{\max}(\alpha'\alpha)$ for a $(p \times r)$-matrix α. The abbreviations $a.s.$, $\mathcal{D}$, $\mathcal{P}$ indicate that results hold almost surely, in distribution or in probability, respectively.

2 Model, analysis and results

In the following the statistical model, the statistical analysis and the asymptotic results are discussed. Initially a Gaussian vector autoregressive model without deterministic components is considered and subsequently generalisation to cases with a constant level or a linear trend and some robustness issues are discussed.

2.1 *A model without deterministic terms*

Consider the statistical model for a p-dimension time series, $X_{-k},\dots,X_0,\dots,X_T$ given by the $(k+1)$-th order autoregressive equation

$$\Delta X_t = \Pi X_{t-1} + \sum_{j=1}^{k} \Gamma_j \Delta X_{t-j} + \varepsilon_t, \tag{2.1}$$

where the initial values, $X_{-k},\dots,X_0$, are fixed and the innovations, $\varepsilon_1,\dots,\varepsilon_T$, are independently, identically normal, $N(0,\Omega)$, distributed. The parameters, $\Pi,\Gamma_1,\dots,\Gamma_k,\Omega$ are $(p\times p)$-matrices and vary freely so that Ω is positive definite.

When Π has reduced rank r it can be written as $\Pi=\alpha\beta'$ for some $(p\times r)$-matrices α,β. The cointegration parameter β can often be associated with long-run economic equilibria since the cumulative effect of the innovations is of less importance for the cointegrating relation $\beta' X_t$ and the differenced process ΔX_t than it is in general for the process X_t. In particular if X_t has $p-r+q$ unit roots then $\beta' X_t$ and ΔX_t have a most q unit roots while the common stochastic trends $\beta_\perp' X_t$ have at least $p-r$ unit roots. The other parameter α or rather $\alpha_\perp$ is associated with the composition of the common stochastic trends. It is therefore of interest to determine the number of cointegrating relations and to test restrictions on α and β.

The hypothesis of at most r cointegrating relations is given by

$$H(r): \qquad \operatorname{rank}\Pi \le r \qquad \text{or} \qquad \Pi = \alpha\beta' \qquad \text{for } \alpha,\beta\in\mathbf{R}^{p\times r}.$$

If $H(r-1)$ is rejected while $H(r)$ is accepted then the cointegration rank is found to be r . Once the rank is determined linear restrictions can be tested on α and β,

$$\begin{aligned} H_\alpha(r): &\qquad \alpha = A\psi, \\ H_\beta(r): &\qquad \beta = H\varphi, \end{aligned}$$

where A,H are known matrices with full column rank and dimensions $(p\times m)$ and $(p\times s)$ respectively.

2.2 *Statistical analysis*

In the presence of the rank restrictions the likelihood function can be analysed using Hotelling's (1936) canonical correlations. For a classical regression model the rank hypothesis was analysed by Bartlett (1938) whereas Anderson (1951) studied the hypothesis $H_\alpha(r)$. The same procedure was introduced in the cointegration context by Johansen (1988, 1996).

The likelihood is maximised in two steps. First, ΔX_t and X_{t-1} are corrected for the remaining terms of equation (2.1) by least squares regression giving the residuals

$$(R_{0,t}, R_{1,t}) = (\Delta X_t, X_{t-1} | \Delta X_{t-1},\dots,\Delta X_{t-k}.) \tag{2.2}$$

Secondly, the squared sample canonical correlations, $1 \geq \hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_p \geq 0$, of $R_{0,t}$ and $R_{1,t}$ are found. This is done by computing the sample product moments

$$\begin{pmatrix} S_{00} & S_{01} \\ S_{10} & S_{11} \end{pmatrix} = \frac{1}{T}\sum_{t=1}^{T}\begin{pmatrix} R_{0,t} \\ R_{1,t} \end{pmatrix}\begin{pmatrix} R_{0,t} \\ R_{1,t} \end{pmatrix}'$$

and then solving either the eigenvalue problem $0 = \det(\lambda S_{11} - S_{10}S_{00}^{-1}S_{01})$ or the "dual" problem $0 = \det(\lambda S_{00} - S_{01}S_{11}^{-1}S_{10})$ for eigenvalues $\hat{\lambda}_j$. The likelihood ratio test statistic for $H(r)$ against the unrestricted model $H(p)$ is then

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

The parameters α, β can be estimated in terms of the eigenvectors associated with the eigenvalue problems. One set of eigenvectors v_j satisfy $\hat{\lambda}_j S_{11}v_j = S_{10}S_{00}^{-1}S_{01}v_j$ and the parameter β is estimated by those corresponding to the r largest eigenvalues and identified through the restriction $\hat{\beta}'S_{11}\hat{\beta} = I_r$ whereas α is estimated by $S_{01}\hat{\beta}(\hat{\beta}'S_{11}\hat{\beta})^{-1}$. Another set of eigenvectors satisfy $\hat{\lambda}_j S_{00}w_j = S_{01}S_{11}^{-1}S_{10}w_j$ and the parameter $\alpha_\perp$ can be estimated by those related to the $p-r$ smallest eigenvalues and identified by $\hat{\alpha}_\perp' S_{00}\hat{\alpha}_\perp = I_{p-r}$.

Some further notation is useful when it comes to analysing the hypotheses on α and β. The residuals R_0 are often pre-multiplied with matrices such as $\overline{\alpha}, \alpha_\perp, \overline{A}$ or $A_\perp$, whereas R_1 is pre-multiplied with $\beta, \beta_\perp, H$ or $H_\perp$. Thus in the notation for the sample product moments "0" and "1" will often be replaced by these symbols, such as $S_{AH} = \overline{A}'S_{01}H$, $S_{AA_\perp} = \overline{A}'S_{00}A_\perp$ and correspondingly $\Omega_{AA_\perp} = \overline{A}'\Omega A_\perp$. The notation $S_{A1\cdot A_\perp} = S_{A1} - S_{AA_\perp}S_{A_\perp A_\perp}^{-1}S_{A_\perp 1}$ is used for partial sample product moments.

Under $H_\beta(r)$ the likelihood is maximised in terms of the squared sample canonical correlations, $1 \geq \hat{\lambda}_1^\beta \geq \cdots \geq \hat{\lambda}_s^\beta \geq 0$, solving $0 = \det(\lambda S_{HH} - S_{H0}S_{00}^{-1}S_{0H})$; see Johansen (1996, Theorem 7.2). The likelihood ratio test statistic for $H_\beta(r)$ against $H(r)$ is

$$LR\left\{ H_\beta\left(r\right) \middle| H\left(r\right)\right\} = T\sum_{j=r+1}^{p}\log\left\{\left(1-\hat{\lambda}_j^\beta\right)\Big/\left(1-\hat{\lambda}_j\right)\right\}.$$

For the analysis of $H_\alpha(r)$ it is convenient to consider the dual eigenvalue problem. Following Johansen (1996, Theorem 8.4 and Section 8.3) the likelihood is maximised in terms of the squared sample canonical correlations, $1 \geq \hat{\lambda}_1^{\alpha,\beta} \geq \cdots \geq \hat{\lambda}_m^{\alpha,\beta} \geq 0$, solving $0 = \det(\lambda S_{AA\cdot A_\perp} - S_{AH\cdot A_\perp}S_{HH\cdot A_\perp}^{-1}S_{HA\cdot A_\perp})$. The likelihood ratio test statistic for $H_\alpha(r)$ against $H_\beta(r)$ is

$$LR\left\{ H_\alpha\left(r\right) \middle| H_\beta\left(r\right)\right\} = T\sum_{j=r+1}^{p}\log\left\{\left(1-\hat{\lambda}_j^{\alpha,\beta}\right)\Big/\left(1-\hat{\lambda}_j^\beta\right)\right\}.$$

2.3 *Representation theorems*

While the autoregressive representation (2.1) is convenient for formulating the statistical analysis an interpretation of the process is provided by its moving average representation. For that purpose the characteristic polynomial of the process is considered. Under the hypothesis $H(r)$ this is

$$A(z) = I_p (1 - z) - \alpha\beta' z - \sum_{j=1}^{k} \Gamma_j z^j (1 - z),$$

and the characteristic roots solve the $\{p(k+1)\}$-th order polynomial equation $0 = \det A(z)$. Under $H(r)$ there are at least $p - r$ unit roots whereas the remaining $pk + r$ roots are the characteristic roots of the vector $(X'_{t-1}\beta, \Delta X'_{t-1}, \ldots, \Delta X'_{t-k})'$. Three conditions are necessary to achieve asymptotic stationarity of $\beta' X_t$ and ΔX_t

(i) $\operatorname{rank} \Pi = r$,

(ii) the number of unit roots, $z = 1$, equals $p - \operatorname{rank}\Pi$,

(iii) the remaining roots are stationary roots, $|z| > 1$.

The condition (ii) can be formulated algebraically as a full rank condition for the $(p-r)$-dimensional square matrix $\alpha'_\perp \Psi \beta_\perp$ where $\Psi = I - \sum_{j=1}^{k} \Gamma_j$. It then follows that the matrix $(\beta, \Psi' \alpha_\perp)$ has full rank.

The Granger Representation can now be formulated.

Theorem 2.1 *(Johansen, 1996, Theorem 4.2, 4.7) Suppose equation (2.1), the hypothesis $H(r)$ and the conditions $(i), (ii), (iii)$ are satisfied. Then $\beta' X_t$ and ΔX_t can be given stationary initial distributions while X_t has the asymptotic representation*

$$X_t = C \sum_{s=1}^{t} \varepsilon_s + \textit{zero mean stationary process} + \tau_c + o_{\mathcal{P}}(1).$$

Here $C = \beta_\perp (\alpha'_\perp \Psi \beta_\perp)^{-1} \alpha'_\perp$ and $\tau_c = C X_0$.

When the conditions (ii) or (iii) are violated the representation is harder to describe as it involves additional types of stochastic components. In empirical analysis the residual processes are often plotted since they are rather stable. This applies even in general cases as indicated in the next two Theorems. The first concerns the "stationary" residuals $\beta' R_{1,t}$ and $R_{0,t}$.

Theorem 2.2 *Suppose equation* (2.1)*, the hypothesis $H(r)$ and condition (i) are satisfied. If in addition either (a) $k > 0$, $\alpha'\Gamma_k = 0$ or (b) $k = 0$, $\overline{\alpha} + \beta = 0$, then*

$$\beta' R_{1,t} = (-\overline{\alpha}' \varepsilon_{t-1} | \Delta X_{t-1}, \ldots, \Delta X_{t-k}), \tag{2.3}$$

$$R_{0,t} = (\varepsilon_t | \Delta X_{t-1}, \ldots, \Delta X_{t-k}). \tag{2.4}$$

The intuition of the condition $\alpha'\Gamma_k = 0$ is that the order of the characteristic polynomial is reduced from $pk + k$ to $pk + p - r$. These roots include $p - r$ unit roots as desired as well as pk other roots which are eliminated by regression on the pk-dimensional vector given by the lagged differences. When adding an extra lag to a well-specified model of order k say, then the additional parameter would be zero $\Gamma_k = 0$ and the condition is trivially satisfied.

The following results describes "common trends" residuals $\beta_\perp' R_{1,t}$. This is related to the representation given by la Cour (1998).

Theorem 2.3 *Suppose equation* (2.1)*, the hypothesis* $H(r)$ *and condition* (i) *are satisfied. Then there exists an* $q > 0$*, dimensions* $s_1, \ldots, s_q \geq 0$ *so* $\sum_{j=1}^q s_j = p - r$ *and* $(p \times s_j)$*-matrices* α_j, β_j *so* $\alpha_j'\alpha_l = \beta_j'\beta_l = 0$ *for* $j \neq l$, $\mathrm{span}(\alpha_1, \ldots, \alpha_q) = \mathrm{span}(\alpha_\perp)$ *and* $\mathrm{span}(\beta_1, \ldots, \beta_q) = \mathrm{span}(\beta_\perp)$ *and in particular*

$$\left(\beta_j' R_{1,t} \middle| \beta' R_{1,t}\right) = \left\{ \frac{-\alpha_j'}{j!} \sum_{s=1}^{t-1} s^{j-1}\varepsilon_s + a_{j,t} \middle| \beta' X_{t-1}, \Delta X_{t-1}, \ldots, \Delta X_{t-k} \right\},$$

with $\max_{t \leq T} ||a_{j,t}|| \overset{a.s.}{=} \mathrm{o}(T^{j-1/2})$. *If in addition condition* (ii) *is satisfied then* $q = 1$.

2.4 *Asymptotic distribution for likelihood ratio test statistics*

For the restrictive situation where the conditions (i), (ii), (iii) are satisfied Johansen (1996) derives the asymptotic distribution of the above mentioned likelihood ratio tests as $T \to \infty$. In the following it will be shown the same asymptotic results hold with less restrictive assumptions.

The first result concerns the squared sample canonical correlations.

Theorem 2.4 *Suppose model* (2.1), *the hypothesis* $H(r)$ *and condition* (i) *are satisfied. Then*

$$\hat{\lambda}_p, \ldots, \hat{\lambda}_{r+1} \overset{a.s.}{\to} 0 \qquad \text{and} \qquad \liminf_{T\to\infty} \hat{\lambda}_r \overset{a.s.}{>} 0.$$

The proof is based on an analysis of the dual eigenvalue problem. In general the argument cannot be based on the other eigenvalue problem since the asymptotic properties of $S_{\beta\beta}$ have only been described in part, see Lemma A.4 in the Appendix.

For the case where the conditions (ii), (iii) are satisfied the result is given by Johansen (1996, Section 11.2). Related results have been found for I(2) models where condition (ii) is violated, see Johansen (1995), and for seasonally integrated systems where (iii) is violated, see Lee (1992). When (ii) and (iii) are indeed satisfied the r largest eigenvalues converge to the solutions of $0 = \det(\lambda\Sigma_{\beta\beta} - \Sigma_{\beta 0}\Sigma_{00}^{-1}\Sigma_{0\beta})$ where $\Sigma_{\beta\beta}, \Sigma_{00}, \Sigma_{\beta 0}$ are the asymptotic covariance matrices for the processes $\beta' X_{t-1}$ and ΔX_t

given past values. A result of this type actually holds whenever $S_{\beta\beta}$ is convergent. It seems reasonable to expect that in the case $\lim_{T\to\infty} S_{\beta\beta}^{-1}$ has reduced rank then the $\{r-\text{rank}(\lim_{T\to\infty} S_{\beta\beta}^{-1})\}$ largest eigenvalues $\hat{\lambda}_j$ converge to one whereas the next $\{\text{rank}(\lim_{T\to\infty} S_{\beta\beta}^{-1})\}$ eigenvalues converge to numbers between zero and one.

The rank test is based on the $(p-r)$ smallest canonical correlations. Their asymptotic behaviour depends on the number of unit roots and hence condition (ii) is needed.

Theorem 2.5 *Suppose model* (2.1), *the hypothesis* $H(r)$ *and the conditions* (*i*), (*ii*) *are satisfied. Let* W *be a* $(p-r)$*-dimension standard Brownian motion. Then*

$$LR\{H(r)|H(p)\} \xrightarrow{\mathcal{D}} \text{tr}\left\{\int_0^1 dW_u W_u' \left(\int_0^1 W_u W_u' du\right)^{-1} \int_0^1 W_u dW_u'\right\}.$$

The proof resembles that of Johansen (1996, Theorem 6.1) who proves the result using condition (iii). For the univariate case a proof is given by Nielsen (2000).

The necessity of the condition (i) is discussed by Johansen (1996, Chapter 12) while the necessity of condition (ii) is observed in I(2) analysis, see Johansen (1995).

A consistent procedure for determining the cointegration rank can be derived from Theorems 2.4 and 2.5. Following Johansen (1996, Chapter 12) the idea is to test $H(0), H(1), \ldots$ sequentially against $H(p)$. The first hypothesis to be accepted determines the rank. If all these tests are conducted using $(1-\delta)$ quantiles of the asymptotic distribution in Theorem 2.5 then the rank estimator $\hat{r}$ is consistent.

Corollary 2.6 *Suppose model* (2.1) *is satisfied. Then*

$$P(\hat{r}=r) \to \begin{cases} 0 & \textit{if } \text{rank}\Pi > r, \\ 1-\delta & \textit{if } \text{rank}\Pi = r \textit{ and condition } (ii) \textit{ is satisfied,} \\ \leq \delta & \textit{if } \text{rank}\Pi < r \textit{ and condition } (ii) \textit{ is satisfied.} \end{cases}$$

The tests on α and β are based on the r largest canonical correlations. A mixed Gaussian argument is used in the asymptotic analysis of the test on β. The mixing parameter is related to $(\beta_\perp' R_{1,t}|\beta' R_{1,t})$ so Theorem 2.3 can be applied.

Theorem 2.7 *Suppose model* (2.1), *the hypothesis* $H_\beta(r)$ *and condition* (*i*) *are satisfied. Then*

$$LR\{H_\beta(r)|H(r)\} \xrightarrow{\mathcal{D}} \chi^2\{r(p-s)\}.$$

When testing restrictions on α then $\beta' R_{1,t}$ appears in the asymptotic argument and Theorem 2.2 with its slightly stronger assumptions is needed. Thus consider a fourth condition

(*iv*) Suppose either (*a*) $k>0$ and $\alpha'\Gamma_k = 0$, (*b*) $k=0$ and $\overline{\alpha}+\beta=0$, or (*c*) the conditions (*ii*), (*iii*) are satisfied.

Theorem 2.8 *Suppose model* (2.1), *the hypothesis* $H_\alpha(r)$ *and the conditions* (*i*), (*iv*) *are satisfied. Then*

$$LR\left\{ H_\alpha\left(r\right)\right| H_\beta\left(r\right)\} \xrightarrow{\mathcal{D}} \chi^2\left\{r\left(p-m\right)\right\}.$$

For the case where also condition (ii), (iii) are satisfied the two results follow from Johansen (1996, Theorem 7.2, 8.2). Johansen (1995) also considers asymptotic theory for β for I(2) models where condition (ii) is violated.

The proofs of Theorems 2.7, 2.8 have two main arguments. First the asymptotic distribution of $\hat{\alpha}_\perp$ is found using the dual eigenvalue problem, see Lemma A.9 in the Appendix. The likelihood ratio test statistics are then written as products of a few likelihood ratio statistics relating the hypotheses to simple hypotheses on $\alpha_\perp$ and each of these are expanded in terms of $(\hat{\alpha}_\perp - \alpha_\perp)$. This is a bit different from Johansen's proof in which the other eigenvalue problem is considered along with simple hypotheses on β.

When testing restrictions on α in the case $k = 0$ the assumption (c) is really necessary. This can be seen from the following example

$$\Delta X_t = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} X_{t-1} + \varepsilon_t. \tag{2.5}$$

The asymptotic distribution of the test for $\alpha' = \psi'(0,1)$ against $H(r)$ is then of the same type as discussed in Theorem 2.5. Compare also with the test for $\alpha'_\perp = (0,1)$ discussed in connection with Lemma B.1 in the Appendix B.

2.5 *A model with a constant term*

For practical purposes the basic model (2.1) would be extended to allow for a non-zero constant level as in the model given by

$$\Delta X_t = (\Pi, \Pi_c)\begin{pmatrix} X_{t-1} \\ 1 \end{pmatrix} + \sum_{j=1}^{k} \Gamma_j \Delta X_{t-j} + \varepsilon_t, \tag{2.6}$$

where Π_c is a p-vector.

Four types of hypotheses are of interest of which the first is the reduced rank hypothesis $H_c(r)$: $\text{rank}(\Pi, \Pi_c) \leq r$ or equivalently $(\Pi, \Pi_c) = \alpha(\beta', \beta'_c)$ for $(p \times r)$-matrices α, β and an r-vector β'_c. Once the rank is determined it may be of interest to test $\beta_c = 0$ so the model reduces to that discussed above $H(r)$: $\text{rank}\Pi \leq r$, $\Pi_c = 0$. When β_c is unrestricted linear restrictions on α, β can be tested using the hypotheses $H^c_\alpha(r)$: $\alpha = A\psi$ and $H^c_\beta(r)$: $\beta = H\varphi$, $\beta'_c \in \mathbf{R}^r$. Test statistics for these hypotheses are derived as in Section 2.2 replacing the residuals (2.2) by

$$\left\{ \Delta X_t, \begin{pmatrix} X_{t-1} \\ 1 \end{pmatrix} \middle| \Delta X_{t-1}, \ldots, \Delta X_{t-k} \right\}. \tag{2.7}$$

For the probabilistic analysis of this model an invariance property is useful.

Theorem 2.9 *Suppose equation* (2.6), *the hypothesis* $H(r)$ *and condition* (i) *are satisfied. Define the p-vector* $\tau_c = -\overline{\beta}\beta_c'$. *Then the process* $X_t^\circ = X_t - \tau_c$ *satisfies* $\beta' X_t^\circ = \beta' X_t' + \beta_c'$, $\beta_\perp' X_t^\circ = \beta_\perp' X_t'$, $\Delta X_t^\circ = \Delta X_t$, *and*

$$\Delta X_t^\circ = \alpha\beta' X_{t-1}^\circ + \sum_{j=1}^{k} \Gamma_j \Delta X_{t-j}^\circ + \varepsilon_t.$$

The canonical correlations of the residuals (2.7) are the same as those of

$$\left\{ \Delta X_t, \left(\begin{array}{c} X_{t-1}|\,1 \\ 1 \end{array} \right) \middle| \Delta X_{t-1}, \ldots, \Delta X_{t-k} \right\}$$

by the invariance of canonical correlations with respect to non-singular linear transformations. When Theorem 2.9 is satisfied it is therefore equivalent to consider the canonical correlations of

$$\left\{ \Delta X_t^\circ, \left(\begin{array}{c} X_{t-1}^\circ \\ 1 \end{array} \right) \middle| \Delta X_{t-1}^\circ, \ldots, \Delta X_{t-k}^\circ \right\}.$$

In other words, for the probabilistic analysis it suffices to assume $\beta_c = 0$ as long as condition (i) is satisfied.

Theorem 2.10 *Suppose model* (2.6), *the hypothesis* $H_c(r)$ *and condition* (i) *are satisfied. Then*

$$\hat{\lambda}_p^c, \ldots, \hat{\lambda}_{r+1}^c \overset{a.s.}{\to} 0, \qquad \text{and} \qquad \liminf_{T\to\infty} \hat{\lambda}_r^c \overset{a.s.}{>} 0.$$

If in addition condition (ii) *is satisfied then*

$$LR\left\{ H_c(r) \middle| H_c(p) \right\} \overset{\mathcal{D}}{\to} tr\left[\int_0^1 dW_u \left(\begin{array}{c} W_u \\ 1 \end{array} \right)' \left\{ \int_0^1 \left(\begin{array}{c} W_u \\ 1 \end{array} \right)^{\otimes 2} du \right\}^{-1} \int_0^1 \left(\begin{array}{c} W_u \\ 1 \end{array} \right) dW_u' \right].$$

If in addition the hypothesis $H(r)$ *and condition* (ii) *are satisfied then*

$$LR\left\{ H(r) \middle| H_c(r) \right\} \overset{\mathcal{D}}{\to} \chi^2(r).$$

If in addition the hypothesis $H_\beta^c(r)$ *is satisfied then*

$$LR\left\{ H_\beta^c(r) \middle| H_c(r) \right\} \overset{\mathcal{D}}{\to} \chi^2\left\{ r(p-s) \right\}.$$

If in addition the hypothesis $H_\alpha^c(r)$ *and condition* (iv) *are satisfied then*

$$LR\left\{ H_\alpha^c(r) \middle| H_c(r) \right\} \overset{\mathcal{D}}{\to} \chi^2\left\{ r(p-m) \right\}$$

The necessity of condition (ii) when testing $H(r)$ against $H_c(r)$ can be checked from the example (2.5) and equation (B.10) in the Appendix.

2.6 *A model with a linear term*

When the data exhibit linear growth it may be of interest to include a linear trend term as in the equation

$$\Delta X_t = (\Pi, \Pi_l)\begin{pmatrix} X_{t-1} \\ t \end{pmatrix} + \mu_c + \sum_{j=1}^{k} \Gamma_j \Delta X_{t-j} + \varepsilon_t. \tag{2.8}$$

In this model a wide range of hypotheses are of interest. The rank hypothesis is as before $H_l(r)$: $\operatorname{rank}(\Pi, \Pi_l) \leq r$ or equivalently $(\Pi, \Pi_l) = \alpha(\beta', \beta_l')$. Once the rank is determined the model is reduced to the previous model by testing $H_c(r)$: $\operatorname{rank}(\Pi, \mu_c) \leq r$ and $\Pi_l = 0$ whereas the linear trend is eliminated from the cointegrating vector by $H_{lc}(r)$: $\beta_l = 0$. Restrictions on α, β are formulated as above as $H_\alpha^l(r)$: $\alpha = A\psi$ and $H_\beta^l(r)$: $\beta = H\varphi$.

This model has weaker invariance properties than the model with a constant.

Theorem 2.11 *Suppose model* (2.8)*, the hypothesis* $H(r)$ *and condition* (i) *are satisfied. If in addition condition* (ii) *is satisfied, or more weakly a* $(p-r)$*-vector* $\overline{\beta}_\perp' \tau_l$ *can be chosen so* $\alpha_\perp' \Psi \beta_\perp \overline{\beta}_\perp' \tau_l = \alpha_\perp'(\mu_c + \Psi\overline{\beta}\delta')$*, then define p-vectors* τ_c, τ_l *so*

$$\beta'\tau_l = -\beta_l', \qquad \alpha_\perp' \Psi \beta_\perp \overline{\beta}_\perp' \tau_l = \alpha_\perp'\left(\mu_c + \Psi\overline{\beta}\delta'\right), \qquad \tau_c = \overline{\beta}\overline{\alpha}'\left(\mu_c - \Psi\tau_l\right).$$

The process $X_{t-1}^\circ = X_{t-1} - \tau_c - \tau_l t$ *then satisfies the equations* $\Delta X_t^\circ = \Delta X_t - \tau_l$, $\beta' X_{t-1}^\circ = \beta_l'(X_{t-1}', t) - \beta'\tau_c$*, and*

$$\Delta X_t^\circ = \alpha\beta' X_{t-1}^\circ + \sum_{j=1}^{k} \Gamma_j \Delta X_{t-j}^\circ + \varepsilon_t.$$

The interpretation of the assumption that $\alpha_\perp' \Psi \beta_\perp \overline{\beta}_\perp' \tau_l = \alpha_\perp'(\mu_c + \Psi\overline{\beta}\delta')$ has a solution is that the possibility of quadratic or higher order polynomial trends is eliminated. This was discussed for an I(2) model with linear trend by Rahbek, Kongsted and Jørgensen (1999). As long as the above assumptions are satisfied the Theorem 2.10 is easily generalised.

The I(2) model with a quadratic trend is studied by Paroulo (1994). For that model the assumptions of the above Theorem 2.11 are violated, but consistency of the eigenvalues and asymptotic results for $\hat{\beta}$ are proved nonetheless. In the proof it is noted that a quadratic trend dominates a cumulated random walk. The process X_t possesses terms of both types but it can be decomposed into components of each type. The same strategy could presumably be applied more generally, but it would be tedious since polynomials and integrated processes of higher order would be involved.

2.7 *Robustness*

The above mentioned asymptotic results remain valid in the more general case of innovations satisfying a Martingale Difference Sequence assumption.

Assumption 2.12 *Let $\mathcal{F}_t$ be an increasing sequence of σ-fields and assume $\{\varepsilon_t, \mathcal{F}_t\}$ is a martingale difference, so $E(\varepsilon_t|\mathcal{F}_{t-1}) \overset{a.s.}{=} 0$ and $E(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}) \overset{a.s.}{=} \Omega$. Further, assume the innovations have bounded conditional moments $\sup_t E(||\varepsilon_t||^{2+\gamma}|\mathcal{F}_{t-1}) \overset{a.s.}{<} \infty$ for some $\gamma > 0$.*

Stronger assumptions are required for the tests on α and β.

Assumption 2.13 *Suppose $\gamma > 4$ in Assumption 2.12 or alternatively that $\gamma > 2$ and condition (iv) is satisfied.*

Theorem 2.14 *Suppose the innovations satisfy Assumption 2.12. Then the asymptotic results given in Theorems 2.3, 2.4, 2.5 and Corollary 2.6 remain valid. If in addition Assumption 2.13 is satisfied then Theorems 2.7, 2.8 remain valid. Theorem 2.10 can be changed in a similar way so Assumption 2.13 is needed when testing restrictions on α, β, β_c.*

The assumption of bounded conditional moments excludes the autoregressive conditional heteroscedasticity process suggested by Engle (1982), that is the univariate process $\varepsilon_t = h_t^{1/2} u_t$ where u_t are independently identically distributed with $Eu_t = 0$, $Eu_t = 1$ and $h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$. When $3\alpha_1^2 < 1$ this process has second moment and $E(\varepsilon_t^2|\mathcal{F}_{t-1}) = h_t$ is time varying and unbounded. While the assumption could be replaced with an unconditional moment condition when proving Functional Central Limit Theorems, see Hansen (1992), it is harder to dispense with it in the proof of equation (C.5) in Appendix C that a process $d_t = Dd_{t-1} + \varepsilon_t$ with $|D| \geq 1$ satisfy $\sum_{t=1}^T |D^{-T} d_{t-1}| \overset{a.s.}{<} \infty$. This fails for the above mentioned process when u_t is Bernoulli distributed on $-1, 1$, $\alpha_0 = 0$ and $|\alpha_1/D^2| > 1$.

The assumption of constant conditional variance would be easier to relax. While it is indeed used by Chan and Wei (1988) it is not used by Lai and Wei (1982, 1983, 1985). For the consistency of the eigenvalues this assumption can therefore be replaced by (a) $T^{-1} \sum_{t=1}^T \varepsilon_t \varepsilon_t' \overset{a.s.}{\to} \Omega$ proved in Lemma A.1, and (b) the limes inferior condition proved in Lemma C.2. To find the distribution of the rank test it would be necessary to revise Lemmas A.10 and A.12 which are based on Chan and Wei (1988). Finally for the tests on α and β it is necessary to make assumptions ensuring that the Lemmas A.2, A.13, A.14 are valid.

APPENDIX: Mathematical Proofs

The appendix is organised as follows. First a series of Lemmas are stated in Section A. Using these results proofs of the main theorems are given in Section B. The Lemmas give convergence results holding with probability one, in probability and in distribution and are proved in the subsequent Sections C-E.

A Some asymptotic results

It is convenient to introduce some notation. The differenced process will be multiplied by the matrix

$$A_T' = M_T^{-1}\begin{pmatrix} \alpha_\perp' \\ \overline{\alpha}' \end{pmatrix}, \qquad \text{where} \qquad M_T = \begin{pmatrix} I_{p-r} & 0 \\ 0 & S_{\beta\beta} \end{pmatrix}^{1/2}.$$

In the analysis of the levels of the process two different transformations are applied. To prove the consistency of the eigenvalues the process is analysed using the basis $(\beta, \beta_\perp)$. The distributional result for the rank test is derived under condition (ii) in which case the matrix $(\beta, \Psi'\alpha_\perp)$ has full rank and is used as basis. Thus define

$$\begin{aligned} B_T' &= M_T^{-1}\begin{pmatrix} I_{p-r} & -\beta_\perp' S_{1\beta}S_{\beta\beta}^{-1} \\ 0 & I_r \end{pmatrix}\begin{pmatrix} \beta_\perp' \\ \beta' \end{pmatrix}, \\ C_T' &= M_T^{-1}\begin{pmatrix} I_{p-r} & -\alpha_\perp'\Psi S_{1\beta}S_{\beta\beta}^{-1} \\ 0 & I_r \end{pmatrix}\begin{pmatrix} \alpha_\perp'\Psi \\ \beta' \end{pmatrix}. \end{aligned}$$

Finally, let $Z_{t-1} = (\Delta X_{t-1}', \ldots, \Delta X_{t-k}')'$.

A.1 *Almost sure results*

The results in this section are based on the work of Lai and Wei (1982, 1983, 1985). Proofs are given in Appendix C. Note, that condition (ii), (iii) are not used.

The following strong law of large numbers is needed

Lemma A.1 *Suppose $\{\varepsilon_t\}$ satisfy Assumption 2.12. Then* $T^{-1}\sum_{t=1}^{T}\varepsilon_t\varepsilon_t' \overset{a.s.}{\to} \Omega$.

A stronger version is used for the distributional analysis of $\hat{\psi}$.

Lemma A.2 *Suppose $\{\varepsilon_t\}$ satisfy Assumption 2.12. Then, for all $\xi < \gamma/(2+\gamma)$ and $\zeta < \min(\xi, 1/2)$*

$$\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t\varepsilon_t' \overset{a.s.}{=} \Omega + \mathrm{o}\left(T^{-\zeta}\right), \qquad \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t \overset{a.s.}{=} \mathrm{o}\left(T^{-\zeta}\right).$$

The order of magnitude of the innovations can be derived from the conditional Borel-Cantelli Theorem.

Lemma A.3 *Suppose* $\{\varepsilon_t\}$ *satisfy Assumption 2.12. Then, for all* $\xi < \gamma/(2+\gamma)$

$$\max_{t \leq T} \|\varepsilon_t\| \overset{a.s.}{=} \mathrm{o}\left\{T^{(1-\xi)/2}\right\}.$$

It has not been possible to describe the distributional behaviour of the cointegrating relations in detail. For the proofs of the main theorems is suffices to show that the limit of the sample covariance matrix is positive definite.

Lemma A.4 *Suppose equation* (2.1)*, the hypothesis* $H(r)$*, condition* (i)*, and Assumption 2.12 are satisfied. Then*

$$\liminf_{T\to\infty} \lambda_{\min}\left(S_{\beta\beta}\right) \overset{a.s.}{>} 0.$$

The asymptotic behaviour of the residual $R_{0,t}$ can be described using the next two Lemmas. The first generalises Lemma A.3 of Nielsen (2000) whereas the second is purely algebraic.

Lemma A.5 *Suppose equation* (2.1)*, the hypothesis* $H(r)$*, condition* (i)*, and Assumption 2.12 are satisfied. Then, for all* $\xi < \gamma/(2+\gamma)$

$$T^{(\xi-1)/2}\left\{\sum_{t=1}^{T}\begin{pmatrix} X_{t-1} \\ Z_{t-1} \\ 1 \\ t \end{pmatrix}^{\otimes 2}\right\}^{-1/2}\sum_{t=1}^{T}\begin{pmatrix} X_{t-1} \\ Z_{t-1} \\ 1 \\ t \end{pmatrix}\varepsilon_t' \overset{a.s.}{\to} 0.$$

Lemma A.6 *Let* (x_t, y_t, a_t) *be a sequence of three vectors and suppose*

$$\left\{\sum_{t=1}^{T}\begin{pmatrix} x_t \\ y_t \end{pmatrix}^{\otimes 2}\right\}^{-1/2}\sum_{t=1}^{T}\begin{pmatrix} x_t \\ y_t \end{pmatrix} a_t' \to 0.$$

Then

$$\left\{\sum_{t=1}^{T}(x_t|y_t)^{\otimes 2}\right\}^{-1/2}\sum_{t=1}^{T}(x_t|y_t)\, a_t' \to 0, \qquad \left\{\sum_{t=1}^{T} y_t^{\otimes 2}\right\}^{-1/2}\sum_{t=1}^{T} y_t a_t' \to 0.$$

Asymptotic results for some of the sample product moment matrices can now be derived. This result generalises Lemma 10.3 of Johansen (1996).

Lemma A.7 *Suppose equation* (2.1), *the hypothesis* $H(r)$, *condition* (i), *and Assumption 2.12 are satisfied. Then, for all* $\xi < \gamma/(2+\gamma)$

$$A_T' S_{00} A_T \overset{a.s.}{=} A_T' \Omega A_T + \begin{pmatrix} 0 & 0 \\ 0 & I_r \end{pmatrix} + \mathrm{o}(1), \qquad (A.1)$$

$$(0, I_r) B_T' S_{10} A_T \overset{a.s.}{\to} (0, I_r), \qquad (A.2)$$

$$\{(0, I_r) B_T' S_{10} A_T - (0, I_r)\} M_T \overset{a.s.}{\to} 0. \qquad (A.3)$$

Suppose $\beta = H\varphi$ *for some* $(p \times s)$ *matrix* H *with full column rank. Then*

$$\begin{aligned} & A_T' S_{0H} S_{HH}^{-1} S_{H0} A_T \\ & \overset{a.s.}{=} \begin{pmatrix} 0 & 0 \\ 0 & I_r \end{pmatrix} + M_T^{-1} \left\{ \begin{pmatrix} 0 \\ I_r \end{pmatrix} S_{\beta\varepsilon}(\alpha_\perp, \overline{\alpha}) + \begin{pmatrix} \alpha_\perp' \\ \overline{\alpha}' \end{pmatrix} S_{\varepsilon\beta}(0, I_r) \right\} M_T^{-1} + \mathrm{o}\left(T^{-\xi}\right), \\ & \overset{a.s.}{=} \begin{pmatrix} 0 & 0 \\ 0 & I_r \end{pmatrix} + \left\{ \begin{array}{cc} o\left(T^{-\xi}\right) & o\left(T^{-\xi/2}\right) \\ o\left(T^{-\xi/2}\right) & o\left(T^{-\xi/2}\right) \end{array} \right\} \end{aligned} \qquad (A.4)$$

and

$$T^{(\xi-1)/2} \sum_{t=1}^{T} \begin{pmatrix} \Delta X_t \\ \varepsilon_t \end{pmatrix} (1|H'X_{t-1}, Z_{t-1}) \left\{ \sum_{t=1}^{T} (1|H'X_{t-1}, Z_{t-1})^2 \right\}^{-1/2} \overset{a.s.}{\to} 0. \qquad (A.5)$$

A Corollary to Lemma A.7 generalising Johansen (1996, Lemma 10.1) is useful.

Corollary A.8 *Suppose equation* (2.1), *the hypothesis* $H(r)$, *condition* (i), *and Assumption 2.12 are satisfied. Let* $\tilde{S}_{00} = A_T' S_{00} A_T$, $\tilde{S}_{01} = A_T' S_{01} B_T$. *Then*

$$\begin{aligned} N &= \tilde{S}_{00}^{-1} - \tilde{S}_{00}^{-1} \tilde{S}_{01} \begin{pmatrix} 0 \\ I_r \end{pmatrix} \left\{ (0, I_r) \tilde{S}_{10} \tilde{S}_{00}^{-1} \tilde{S}_{01} \begin{pmatrix} 0 \\ I_r \end{pmatrix} \right\}^{-1} (0, I_r) \tilde{S}_{10} \tilde{S}_{00}^{-1} \\ & \overset{a.s.}{\to} \begin{pmatrix} I_{p-r} \\ 0 \end{pmatrix} (\alpha_\perp' \Omega \alpha_\perp)^{-1} (I_{p-r}, 0). \end{aligned}$$

The consistency of $\hat{\alpha}_\perp$ can now be stated. Suppose $\alpha = A\psi$ and $\beta = H\varphi$. For some random $\{r \times (m-r)\}$-matrix U_T let

$$\tilde{\psi}_\perp = \hat{\psi}_\perp \left(\overline{\psi}_\perp' \hat{\psi}_\perp \right)^{-1} = \psi_\perp + \overline{\psi} \psi' \tilde{\psi}_\perp \overset{\text{def}}{=} \psi_\perp + \overline{\psi} U_T. \qquad (A.6)$$

Lemma A.9 *Suppose equation* (2.1), *the hypotheses* $H_\alpha(r)$, $H_\beta(r)$, *condition* (i), *and Assumption* 2.12 *are satisfied. Let* $\overline{A}_\omega = \overline{A} - A_\perp \Omega_{A_\perp A_\perp}^{-1} \Omega_{A_\perp A}$. *Then, for all* $\xi < \gamma/(2+\gamma)$

$$S_{\beta\beta}^{1/2} U_T \overset{a.s.}{=} \mathrm{o}\left(T^{-\xi/2}\right), \qquad (A.7)$$

$$\tilde{\psi}_\perp \overset{a.s.}{=} \psi_\perp + \mathrm{o}\left(T^{-\xi/2}\right), \qquad (A.8)$$

$$-S_{\beta\beta}^{1/2} U_T \overset{a.s.}{=} S_{\beta\beta}^{-1/2} S_{\beta\varepsilon} \overline{A}_\omega \psi_\perp \left\{ 1 + \mathrm{o}\left(T^{-\xi/2}\right) \right\} + \mathrm{o}\left(T^{-\xi}\right). \qquad (A.9)$$

A.2 *A convergence in probability result*

The asymptotic behaviour of the residuals $R_{1,t}$ can be described by generalising Lemmas A.4 and B.1 of Nielsen (2000). It suffices to prove convergence in probability since the common trends are not normalised by $S_{\beta\beta}$.

Lemma A.10 *Suppose equation* (2.1)*, the hypothesis, $H(r)$, the conditions* (*i*), (*ii*), *and Assumption 2.12 are satisfied. Then, for all $\eta > 0$ and $k = 0, 1, \ldots$*

$$T^{-\eta-k}\left\{\sum_{t=1}^{T}\begin{pmatrix}\beta' X_{t-1}\\ Z_{t-1}\end{pmatrix}^{\otimes 2}\right\}^{-1/2}\sum_{t=1}^{T}\begin{pmatrix}\beta' X_{t-1}\\ Z_{t-1}\end{pmatrix}t^{k}\xrightarrow{\mathcal{P}}0,$$

$$T^{-\eta-k-1/2}\left\{\sum_{t=1}^{T}\begin{pmatrix}\beta' X_{t-1}\\ Z_{t-1}\end{pmatrix}^{\otimes 2}\right\}^{-1/2}\sum_{t=1}^{T}\begin{pmatrix}\beta' X_{t-1}\\ Z_{t-1}\end{pmatrix}\sum_{s=1}^{t-1}(t-s)^{k}\varepsilon_s\xrightarrow{\mathcal{P}}0.$$

If condition (*ii*) *is not satisfied the results only hold for $\eta > 1/2$.*

A.3 *Distributional results*

The basic distributional result follows from Chan and Wei (1988, Theorem 2.2).

Lemma A.11 *Suppose $\{\varepsilon_t\}$ satisfy Assumption 2.12. Let W be a p-dimensional Brownian motion with variance Ω. Then, for $u \in [0, 1]$,*

$$\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{[Tu]}\varepsilon_s, \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{t-1}\varepsilon_s\varepsilon_t'\right)\xrightarrow{\mathcal{D}}\left(W_u, \int_0^1 W_v dW_v'\right)$$

on $D[0,1]^p \times \mathbf{R}^{p\times p}$, where $D[0,1]$ is the space of functions on $[0,1]$ which are right-continuous and have left-limits.

The distributional behaviour of the common trend components of the product moment matrices can now be described.

Lemma A.12 *Suppose equation* (2.1), *the hypothesis $H(r)$, the conditions* (*i*), (*ii*), *and Assumption 2.12 are satisfied. Then*

$$\left\{(I_{p-r}, 0)\, C_T' S_{10} A_T M_T, \begin{pmatrix}T^{-1/2} I_{p-r} & 0\\ 0 & I_r\end{pmatrix} C_T' S_{11} C_T \begin{pmatrix}T^{-1/2} I_{p-r} & 0\\ 0 & I_r\end{pmatrix}\right\}$$
$$\xrightarrow{\mathcal{D}}\left\{\alpha_\perp'\int_0^1 W_u dW_u'\,(\alpha_\perp, \overline{\alpha}), \begin{pmatrix}\alpha_\perp'\int_0^1 W_u W_u' du\,\alpha_\perp & 0\\ 0 & I_r\end{pmatrix}\right\}.$$

For the test for $H_\alpha(r)$ the following version of the Brown's (1971) Martingale Central Limit Theorem is used.

Lemma A.13 *Suppose equation* (2.1), *the hypothesis* $H_\alpha(r)$, *condition* (i), *Assumption 2.12 and* $E||\varepsilon_t||^4 < \infty$ *are satisfied. Then*

$$T^{-1/2}\Omega_{\alpha_\perp\alpha_\perp}^{-1/2}\alpha_\perp'\sum_{t=1}^{T}\varepsilon_t\varepsilon_{t-1}'\overline{\alpha}\Omega_{\alpha\alpha}^{-1/2}\xrightarrow{\mathcal{D}} N\left\{0, I_{r(p-r)}\right\},$$

Finally, for the test for $H_\beta(r)$ a mixed Gaussian result is useful. Let $\omega = \Omega_{\alpha\alpha_\perp}\Omega_{\alpha_\perp\alpha_\perp}^{-1}$ and $\overline{\alpha}_\omega' = \overline{\alpha}' - \omega\alpha_\perp'$ so $\overline{\alpha}_\omega'\varepsilon_t$ and $(R_{\beta_\perp,t}|R_{\beta,t})$ are approximately uncorrelated by Theorem 2.3.

Lemma A.14 *Suppose equation* (2.1), *the hypothesis* $H_\beta(r)$, *condition* (i) *and Assumption 2.12 are satisfied. Then*

$$\Omega_{\alpha\alpha\cdot\alpha_\perp}^{-1/2}\overline{\alpha}_\omega' S_{\varepsilon H_\perp\cdot H} S_{H_\perp H_\perp\cdot H}^{-1/2}\xrightarrow{\mathcal{D}} N\left\{0, I_{r(p-s)}\right\}.$$

B Proofs of main Theorems

The main theorems are proved under the assumptions stated in Theorem 2.14.

B.1 *Representation Theorems*

Proof of Theorem 2.2. (2.3): Rearrange the model equation (2.1), pre-multiply by $\overline{\alpha}'$ to see that

$$\beta' X_t = -\overline{\alpha}'\varepsilon_t + (\overline{\alpha}' + \beta')\,\Delta X_t - \sum_{j=1}^{k} \overline{\alpha}'\Gamma_j \Delta X_{t-j}.$$

and exploit the regression on $\Delta X_{t-1}, \ldots, \Delta X_{t-k}$.

(2.4): The result for $R_{0,t}$ follows from the identity $R_{0,t} = \alpha\beta' R_{1,t} + R_{\varepsilon,t}$ ■

Proof of Theorem 2.3. An iterative argument is used. In the first step use that $\Delta X_{t-1} - \Delta X_{t-1-j} = \sum_{l=1}^{j} \Delta^2 X_{t-l}$ and rearrange equation (2.1) as

$$\alpha\beta' X_{t-1} = \left(I_p + \alpha\beta' - \sum_{j=1}^{k}\Gamma_j\right)\Delta X_{t-1} + \sum_{j=1}^{k}\left(\sum_{l=j}^{k}\Gamma_l\right)\Delta^2 X_{t-j} - \varepsilon_{t-1}. \tag{B.1}$$

Introduce the notation $s_0 = r$, $\beta_0 = \beta$, $\beta_{0\perp} = \beta_\perp$, $\alpha_0 = \alpha$, $\alpha_{0\perp} = \alpha_\perp$, $C_{00} = I_p + \alpha\beta' - \sum_{j=1}^{k}\Gamma_j$ and $C_{0j} = \sum_{l=j}^{k}\Gamma_l$. Pre-multiplying (B.1) by $(\overline{\alpha}_0, \alpha_{0,\perp})'$ gives

$$\begin{pmatrix} \beta_0' X_{t-1} \\ 0 \end{pmatrix} = \begin{pmatrix} \overline{\alpha}_0' \\ \alpha_{0\perp}' \end{pmatrix}\left(C_{00}\Delta X_{t-1} + \sum_{j=1}^{k} C_{0j}\Delta^2 X_{t-j} - \varepsilon_{t-1}\right). \tag{B.2}$$

To get to the n-th step define a function $\kappa_{s,n}$ as $1_{(s=t-1)}$ for $n = 0$ and $s^{n-1}/n!$ for $n > 0$, and let $\delta_{n,t}$ be a function identical to zero for $n = 0$ and satisfying $\max_{t\le T}||\delta_{n,t}|| \overset{a.s.}{=} \mathrm{o}(T^{n-1/2-\xi/2})$ for some $\xi < \gamma/(2+\gamma)$ and $n > 0$. Then, with $n = 0$, equation (B.2) implies that

$$0 = \alpha_{n\perp}'\left(C_{n0}\Delta X_{t-1} + \sum_{j=1}^{k} C_{nj}\Delta^2 X_{t-j} - \sum_{s=1}^{t-1}\kappa_{s,n}\varepsilon_s + \delta_{n,t}\right). \tag{B.3}$$

A recursion formula for C_{n0} is given by la Cour (1998). She uses the notation $C_{n0} = F^n(A)_1(1)$ and parametrise Π as $\Pi = -\alpha\beta'$. In (B.3) use the identity $I_p = \sum_{j=0}^{n}\overline{\beta}_j\beta_j' + \overline{\beta}_{n\perp}\beta_{n\perp}'$ and sum over t so see that

$$-\left(\alpha_{n\perp}' C_{n0}\overline{\beta}_{n\perp}\right)\beta_{n\perp}' X_{t-1} = \tag{B.4}$$

$$\alpha_{n\perp}'\left(C_{n0}\sum_{j=0}^{n}\overline{\beta}_j\beta_j' X_{t-1} + \sum_{j=1}^{k} C_{nj}\Delta X_{t-j} - \sum_{s=1}^{t-1}\kappa_{s,n+1}\varepsilon_s + \sum_{s=1}^{t-1}\delta_{n,s} + A_n\right).$$

For $n=0$ then $\sum_{s=1}^{t-1}\delta_{0,s}=0$ and A_0 is a linear combination of the initial values and thus of order $\mathrm{O}(1)$. For $n>0$ then $\max_{t\leq T}||\sum_{s=1}^{t-1}\delta_{n,s}||\overset{a.s.}{=}\mathrm{o}(T^{n+1/2-\xi/2})$, whereas A_n involves a term depending on initial values and is of order $\mathrm{O}(1)$. The boundedness of the innovations stated in Lemma A.3 shows that $A_n\overset{a.s.}{=}\mathrm{o}(T^{n+1/2-\xi/2})$.

Suppose the $(p-\sum_{j=0}^{n}s_j)$-dimensional square matrix $\alpha_{n\perp}'C_{n0}\overline{\beta}_{n\perp}$ has rank $s_{n+1}\leq(p-\sum_{j=0}^{n}s_j)$. Then it can be written as $\zeta_n\eta_n'$ for some $\{(p-\sum_{j=0}^{n}s_j)\times s_{n+1}\}$-dimensional matrices ζ_n,η_n with full column rank. Pre-multiplying equation (B.4) by $(\overline{\zeta}_n,\zeta_{n\perp})'$ then gives

$$\begin{pmatrix}-\eta_n'\beta_{n\perp}'X_{t-1}\\0\end{pmatrix}= \tag{B.5}$$

$$\begin{pmatrix}\overline{\zeta}_n'\\\zeta_{n\perp}'\end{pmatrix}\alpha_{n\perp}'\left(C_{n0}\sum_{j=0}^{n}\overline{\beta}_j\beta_j'X_{t-1}+\sum_{j=1}^{k}C_{nj}\Delta X_{t-j}-\sum_{s=1}^{t-1}\kappa_{s,n+1}\varepsilon_s+\sum_{s=1}^{t-1}\delta_{n,s}+A_n\right).$$

Introduce $(p\times s_{n+1})$-matrices $\alpha_{n+1}=\alpha_{n\perp}\overline{\zeta}_n$, $\beta_{n+1}=-\beta_{n\perp}\eta_n$, and $\{p\times(p-\sum_{j=0}^{n+1}s_j)\}$-matrices $\alpha_{n+1,\perp}=\alpha_{n\perp}\zeta_{n\perp}$, $\beta_{n+1,\perp}=\beta_{n\perp}\eta_{n\perp}$.

If $\sum_{j=0}^{n+1}s_j<p$ then in the $\alpha_{n+1,\perp}$-component of equation (B.5), substitute $\beta_j'X_{t-1}$, first for $j=n,\ldots,1$ using the α_{n+1}-component and then for $j=0$ using (B.2). This results in an expression of the form (B.3) where $\delta_{n+1,t}$ is a linear combination of A_j, $\sum_{s=1}^{t-1}\delta_{j,s}$ and $\sum_{s=1}^{t-1}\kappa_{s,j}$ for $j=1,\ldots,n$. By Lemma A.3 then $\max_{t\leq T}||\delta_{n+1,t}||=\mathrm{o}(T^{n+1/2-\xi/2})$ as required above. The iterative step is repeated for $n+1$.

When $\sum_{j=0}^{n+1}s_j=p$ the iteration stops and the $\alpha_{n+1,\perp}$-component in (B.5) vanishes. The α_{n+1}-component shows

$$\left(\beta_{j+1}'X_{t-1}\middle|\beta'X_{t-1},Z_{t-1}\right)=\left(\alpha_{j+1}'\sum_{s=1}^{t-1}\kappa_{s,j+1}\varepsilon_s+a_{j+1,t}\middle|\beta'X_{t-1},Z_{t-1}\right)$$

for $0\leq j<n$ with $a_{j+1,t}=\alpha_{j+1}'(C_{n0}\sum_{l=1}^{j}\overline{\beta}_l\beta_l'X_{t-1}+\sum_{s=1}^{t-1}\delta_{j,s}+A_j)$. A sequential argument for $j=0,\ldots,n$ shows that $\max_{t\leq T}||a_{j+1,t}||=\mathrm{o}(T^{j-1/2})$.

Note that the iteration eventually comes to an end. Analysis of the characteristic polynomial $A(z)$ shows that $p-\sum_{j=1}^{n+1}s_j$ unit roots are identified in each step and the process only has a finite number of characteristic roots. ∎

B.2 *The rank test*

Proof of Theorem 2.4. The solutions to the dual eigenvalue problem, $0=\det(\lambda S_{00}-S_{01}S_{11}^{-1}S_{10})$, equal those of $0=\det A_T'\left\{\lambda S_{00}-S_{01}S_{11}^{-1}S_{10}\right\}A_T$. By (A.1), (A.4) of Lemma A.7 it is equivalent to consider $0\overset{a.s.}{=}\det\{P(\lambda)+\mathrm{o}(1)\}$ where

$$P(\lambda)=\lambda M_T^{-1}\begin{pmatrix}\alpha_\perp'\\\overline{\alpha}'\end{pmatrix}\Omega(\alpha_\perp,\overline{\alpha})M_T^{-1}+(\lambda-1)\begin{pmatrix}0&0\\0&I_r\end{pmatrix}.$$

For $\lambda = 0$ the matrix polynomial $P(\lambda)$ satisfies $(I_{p-r}, 0)P(\lambda) = \lambda(I_{p-r}, 0)$ and therefore the $p - r$ smallest eigenvalues converge to zero. The equation $0 = \det P(\lambda)$ can be rewritten as $0 = \det\left(\lambda\alpha_\perp'\Omega\alpha_\perp\right)\det\left\{(\lambda-1)\,I_r + \lambda S_{\beta\beta}^{-1/2}\Omega_{\alpha\alpha\cdot\alpha_\perp}S_{\beta\beta}^{-1/2}\right\}$. Since $S_{\beta\beta}^{-1/2} \overset{a.s.}{<} \infty$ by Lemma A.4 it follows that $\liminf_{T\to\infty}\hat{\lambda}_r \overset{a.s.}{>} 0$. ■

Proof of Theorem 2.5. In order to analyse the smallest $p - r$ eigenvalues replace λ by λ/T in the eigenvalue problem. By multiplying with C_T this becomes $0 = \det C_T'\left(T^{-1}\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}\right)C_T$, which is asymptotically equivalent to

$$0 = \det\left\{\frac{\lambda}{T}\begin{pmatrix} I_{p-r} & 0 \\ 0 & 0\end{pmatrix} C_T'S_{11}C_T\begin{pmatrix} I_{p-r} & 0 \\ 0 & 0\end{pmatrix} - C_T'S_{10}S_{00}^{-1}S_{01}C_T\right\}.$$

Using the matrix N defined in Corollary A.8 this determinant can be rewritten as

$$\begin{aligned} 0 \;=\; & \det\left\{-\left(0, I_r\right)C_T'S_{10}S_{00}^{-1}S_{01}C_T\begin{pmatrix} 0 \\ I_r\end{pmatrix}\right\} \\ & \times\det\left\{\left(I_{p-r}, 0\right)C_T'\left(\frac{\lambda}{T}S_{11} - S_{10}A_TNA_TS_{01}\right)C_T\begin{pmatrix} 0 \\ I_r\end{pmatrix}\right\}\end{aligned}$$

The first term does not depend on λ. Thus, using Corollary A.8 the asymptotic distribution of the $p - r$ smallest eigenvalues can be found from

$$0 = \det\left[\left(I_{p-r}, 0\right)C_T'\left\{\frac{\lambda}{T}S_{11} - S_{10}A_T\begin{pmatrix} I_{p-r} \\ 0\end{pmatrix}\Omega_{\alpha_\perp\alpha_\perp}^{-1}\left(I_{p-r}, 0\right)A_TS_{01}\right\}C_T\begin{pmatrix} 0 \\ I_r\end{pmatrix}\right]$$

Lemma A.12 together with the Continuous Mapping Theorem, see Billingsley (1968), shows that this expression is asymptotically equivalent to

$$0 = \det\alpha_\perp'\left\{\lambda\int_0^1 W_uW_u'du - \int_0^1 W_udW_u'\alpha_\perp\left(\alpha_\perp'\Omega\alpha_\perp\right)^{-1}\alpha_\perp'\int_0^1 dW_uW_u'\right\}\alpha_\perp,$$

where $(\alpha_\perp'\Omega\alpha_\perp)^{-1/2}\alpha_\perp'W$ is a standard Brownian motion. ■

B.3 *Test for simple hypothesis on* $\alpha_\perp$ *when* $\beta = H\varphi$ *and* $\alpha = A\psi$

In the following the likelihood ratio test for a simple hypothesis on $\alpha_\perp$ is considered when both $H_\alpha(r)$: $\alpha = A\psi$ and $H_\beta(r)$: $\beta = H\varphi$ are satisfied. That is, the hypothesis $H_{\alpha_\perp^\circ}(r)$: $\alpha_\perp = \left(A_\perp, \overline{A}\psi_\perp^\circ\right)$ for some known $\psi_\perp^\circ$ is considered.

Lemma B.1 *Suppose equation* (2.1), $\alpha_\perp = (A_\perp, \overline{A}\psi_\perp^\circ)$, *the hypotheses* $H_\alpha(r)$, $H_\beta(r)$, *condition* (i), *and Assumption 2.13 are satisfied. Then*

$$LR\left\{H_{\alpha_\perp}(r)\middle|\,H_\alpha(r), H_\beta(r)\right\} = \operatorname{tr}\left\{\left(\psi_\perp'\Omega_{AA\cdot A_\perp}\psi_\perp\right)^{-1}T\psi_\perp'\overline{A}_\omega'S_{\varepsilon\beta}S_{\beta\beta}^{-1}S_{\beta\varepsilon}\overline{A}_\omega\psi_\perp\right\} + o_{\mathcal{P}}(1).$$

Proof of Lemma B.1. The model equation (2.1) concentrated with respect to the lagged differences is given by $R_{0,t} = A\psi\varphi' H' R_{1,t} + \hat{\varepsilon}_t$. Pre-multiplication by $(\overline{A}, A_\perp)'$ shows that $A_\perp' X_t$ is weakly exogenous for ψ, φ. Thus likelihood inference can be based on the partial system of $\overline{A}' X_t$ given $A_\perp' X_t$,

$$\overline{A}' R_{0,t} = \psi\varphi' H' R_{1,t} + \Omega_{AA_\perp}\Omega_{A_\perp A_\perp}^{-1} A_\perp' R_{0,t} + \overline{A}_\omega \hat{\varepsilon}_t.$$

Following Johansen (1996, p. 130) the concentrated likelihood function of $\psi_\perp$ is

$$L^{-2/T}(\psi_\perp) = \max_{\varphi,\Omega,\Gamma_j} L^{-2/T}(\psi_\perp, \varphi, \Omega, \Gamma_j) = |S_{AA\cdot H,A_\perp}| \frac{|\psi_\perp' S_{AA\cdot A_\perp}\psi_\perp|}{|\psi_\perp' S_{AA\cdot H,A_\perp}\psi_\perp|}.$$

When $\hat{\psi}_\perp$ denotes the unrestricted maximum likelihood estimator of $\psi_\perp$ the likelihood ratio test statistic can be expressed as

$$LR\left\{ H_{\alpha_\perp^\circ}(r) \middle| H_\alpha(r), H_\beta(r) \right\} = T\left\{ \log \frac{|\psi_\perp^{\circ\prime} S_{AA\cdot A_\perp}\psi_\perp^\circ|}{|\psi_\perp^{\circ\prime} S_{AA\cdot H,A_\perp}\psi_\perp^\circ|} - \log \frac{\left|\hat{\psi}_\perp' S_{AA\cdot A_\perp}\hat{\psi}_\perp\right|}{\left|\hat{\psi}_\perp' S_{AA\cdot H,A_\perp}\hat{\psi}_\perp\right|} \right\}.$$

Without changing the value of the test statistic $\hat{\psi}_\perp$ can be replaced by $\tilde{\psi}_\perp = \hat{\psi}_\perp(\overline{\psi}_\perp' \hat{\psi}_\perp)^{-1}$. Following Johansen (1996, Lemma A.8) an expansion around $\psi_\perp = \psi_\perp^\circ$ gives

$$LR\left\{ H_{\alpha_\perp^\circ}(r) \middle| H_\alpha(r), H_\beta(r) \right\} = T\mathrm{tr}(D_1 - D_2) + \mathrm{O}\left(\left\|\tilde{\psi}_\perp - \psi_\perp\right\|^3 \right).$$

where $\tilde{\psi}_\perp - \psi_\perp \overset{a.s.}{\to} 0$ by (A.8) in Lemma A.9 and

$$\begin{aligned} D_1 &= (\psi_\perp' S_{AA\cdot A_\perp}\psi_\perp)^{-1} \left(\tilde{\psi}_\perp - \psi_\perp\right)' S_{AA\cdot A_\perp,\overline{A}\psi_\perp} \left(\tilde{\psi}_\perp - \psi_\perp\right), \\ D_2 &= (\psi_\perp' S_{AA\cdot H,A_\perp}\psi_\perp)^{-1} \left(\tilde{\psi}_\perp - \psi_\perp\right)' S_{AA\cdot H,A_\perp,\overline{A}\psi_\perp} \left(\tilde{\psi}_\perp - \psi_\perp\right). \end{aligned}$$

Using Lemma A.2 and (A.4) of Lemma A.7 it is seen that $\psi_\perp' S_{AA\cdot A_\perp}\psi_\perp \overset{a.s.}{=} \psi_\perp' \Omega_{AA\cdot A_\perp}\psi_\perp + \mathrm{o}(T^{-\xi/2})$ and $\psi_\perp' S_{AA\cdot H,A_\perp}\psi_\perp \overset{a.s.}{=} \psi_\perp' \Omega_{AA\cdot A_\perp}\psi_\perp + \mathrm{o}(T^{-\xi/2})$. Applying the definition $\tilde{\psi}_\perp - \psi_\perp = \overline{\psi} U_T$ it follows

$$D_1 - D_2 \overset{a.s.}{=} (\psi_\perp' \Omega_{AA\cdot A_\perp}\psi_\perp)^{-1} U_T' \overline{\psi}' (S_{AA\cdot\alpha_\perp} - S_{AA\cdot H,\alpha_\perp,}) \overline{\psi} U_T \left\{ 1 + \mathrm{o}\left(T^{-\xi/2}\right) \right\}.$$

Equation (A.4) shows that $S_{HH\cdot\alpha_\perp} \overset{a.s.}{=} S_{HH}\{1 + \mathrm{o}(T^{-\xi})\}$ while (A.1), (A.4) and Lemmas A.4, A.5, A.6 imply $S_{\beta\beta}^{-1/2}\overline{\psi}' S_{AH\cdot\alpha_\perp} S_{HH}^{-1} S_{HA\cdot\alpha_\perp}\overline{\psi} S_{\beta\beta}^{-1/2} \overset{a.s.}{=} I_r + \mathrm{o}(T^{-\xi/2})$ and therefore

$$\begin{aligned} T(D_1 - D_2) &\overset{a.s.}{=} (\psi_\perp' \Omega_{AA\cdot A_\perp}\psi_\perp)^{-1} T U_T' S_{\beta\beta} U_T \left\{ 1 + \mathrm{o}\left(T^{-\xi/2}\right) \right\} \\ &\overset{a.s.}{=} T\psi_\perp' \overline{A}_\omega' S_{\varepsilon\beta} S_{\beta\beta}^{-1} S_{\beta\varepsilon} \overline{A}_\omega \psi_\perp \left\{ 1 + \mathrm{o}\left(T^{-\xi/2}\right) \right\} + \mathrm{o}\left(T^{1/2-\xi} \psi_\perp' \overline{A}_\omega' S_{\varepsilon\beta} S_{\beta\beta}^{-1/2} \right), \end{aligned}$$

where equation (A.9) of Lemma A.9 has been used.

If $\gamma > 2$ in Assumption 2.13 and either of the conditions (iv, a) or (iv, b) is satisfied then $\psi_\perp' \overline{A}_\omega' S_{\varepsilon\beta} S_{\beta\beta}^{-1/2} = \mathrm{O}_\mathcal{P}(T^{-1/2})$ by Theorem 2.2 and Lemma A.14.

If $\gamma > 2$ and condition (iv, c) is satisfied then $\beta' X_t$ is asymptotically stationary by Theorem 2.1 and the Central Limit Theorem for linear processes with exponentially decreasing coefficients, see Phillips and Solo (1992), implies $\psi_\perp' \overline{A}_\omega' S_{\varepsilon\beta} S_{\beta\beta}^{-1/2} = \mathrm{O}_\mathcal{P}(T^{-1/2})$.

If $\gamma > 4$ then ξ can be chosen so $\xi > 2/3$ and the result follows since $S_{\varepsilon\beta} S_{\beta\beta}^{-1/2} \overset{a.s.}{=} \mathrm{o}(T^{-\xi/2})$ by the Lemmas A.5, A.6. ■

B.4 *Test for* $\alpha = A\psi$ *when* $\beta = H\varphi$

Proof of Theorem 2.8. If condition (iv,c) is satisfied the result follows from Johansen (1996, Chapter 8). Otherwise, if (iv,a) or (iv,b) is satisfied let $L(\alpha_\perp, \varphi, \Omega)$ denote the likelihood concentrated with respect to the lagged differences, under $H_\beta(r)$: $\beta = H\varphi$. The hypothesis $H_\alpha(r)$ can be formulated as $\alpha_\perp = (A_\perp, \overline{A}\psi_\perp)$. Thus, the likelihood ratio test statistic can be rewritten as

$$Q\left\{ H_\alpha(r) | H_\beta(r) \right\} = \max_{\psi_\perp, \varphi, \Omega} L\left(A_\perp, \overline{A}\psi_\perp, \varphi, \Omega\right) \Big/ \max_{\alpha_\perp, \varphi, \Omega} L\left(\alpha_\perp, \varphi, \Omega\right).$$

Extend this fraction by the maximised likelihood when $\alpha_\perp^\circ = (A_\perp, \overline{A}\psi_\perp^\circ)$ is known

$$Q\left\{ H_\alpha(r) | H_\beta(r) \right\} = \left\{ \frac{\max_{\psi_\perp, \varphi, \Omega} L\left(A_\perp, \overline{A}\psi_\perp, \varphi, \Omega\right)}{\max_{\varphi, \Omega} L\left(A_\perp, \overline{A}\psi_\perp^\circ, \varphi, \Omega\right)} \right\} \left\{ \frac{\max_{\varphi, \Omega} L\left(\alpha_\perp^\circ, \varphi, \Omega\right)}{\max_{\alpha_\perp, \varphi, \Omega} L\left(\alpha_\perp, \varphi, \Omega\right)} \right\}.$$

Combine the result (2.3) in Theorem 2.2 with the Lemmas A.5, B.1 to see that

$$\begin{aligned} & LR\left\{ H_{\alpha_\perp^\circ}(r) \Big| H_\alpha(r), H_\beta(r) \right\} \\ & \overset{a.s.}{=} \operatorname{tr}\left\{ \left(\psi_\perp' \Omega_{AA \cdot A_\perp} \psi_\perp\right)^{-1/2} T^{-1/2} \psi_\perp' \overline{A}_\omega' \sum_{t=1}^T \varepsilon_t \varepsilon_{t-1}' \overline{\alpha} \Omega_{\alpha\alpha}^{-1/2} \right\}^{\otimes 2} \left\{1 + \mathrm{o}(1)\right\} + \mathrm{o}(1). \end{aligned}$$

Noting that $\alpha_\perp(\alpha_\perp' \Omega \alpha_\perp)^{-1} \alpha_\perp = \overline{A}_\omega \psi_\perp (\psi_\perp' \Omega_{AA \cdot A_\perp} \psi_\perp)^{-1} \psi_\perp' \overline{A}_\omega' + A_\perp (A_\perp' \Omega A_\perp)^{-1} A_\perp'$ it follows that

$$LR\left\{ H_\alpha(r) | H_\beta(r) \right\} \overset{a.s.}{=} \operatorname{tr}\left\{ \left(A_\perp' \Omega A_\perp\right)^{-1/2} T^{-1/2} A_\perp' \sum_{t=1}^T \varepsilon_t \varepsilon_{t-1}' \overline{\alpha} \Omega_{\alpha\alpha}^{-1/2} \right\}^{\otimes 2} + \mathrm{o}(1)$$

The Martingale Central Limit Theorem A.13 shows that this is asymptotically χ^2 distributed with $\dim(\alpha' \varepsilon_t) \dim(A_\perp' \varepsilon_t) = r(p - m)$ degrees of freedom. ■

B.5 *Test for* $\beta = H\varphi$ *when* $\alpha_\perp$ *is known*

When $\alpha_\perp = A_\perp$ is known then $\alpha = A\psi$ for some r-dimensional square matrix ψ. It can be assumed without loss of generality that $\psi = I_r$ and hence the concentrated model equation can be written as $R_{0,t} = A\varphi' H' R_{1,t} + \hat{\varepsilon}_t$. Since $\alpha_\perp = A_\perp$ is known then $\alpha'_\perp R_{0,t}$ is weakly exogenous for β and likelihood inference about β can be performed in the partial system

$$\overline{\alpha}' R_{0,t} = \beta' R_{1,t} + \omega \alpha'_\perp R_{0,t} + \overline{\alpha}'_\omega \hat{\varepsilon}_t. \tag{B.6}$$

The hypothesis H_β can therefore be analysed using standard regression methods so

$$\begin{aligned} LR\left\{ H_\beta(r) \middle| H_{\alpha_\perp^\circ}(r) \right\} &= -T \log \left(S_{\alpha\alpha\cdot 1,\alpha_\perp} S^{-1}_{\alpha\alpha\cdot H,\alpha_\perp} \right) \\ &= -T \log \det \left(I_r - S^{-1}_{\alpha\alpha\cdot H,\alpha_\perp} S_{\alpha H_\perp \cdot H,\alpha_\perp} S^{-1}_{H_\perp H_\perp \cdot H,\alpha_\perp} S_{H_\perp \alpha \cdot H,\alpha_\perp} \right), \end{aligned}$$

where it has been used that regression on $R_{1,t}$ is equivalent to regression on $(H' R_{1,t}, H'_\perp R_{1,t})$. The asymptotic distribution of this statistic is given as

Lemma B.2 *Suppose equation* (2.1)*, the hypothesis* $H_\beta(r)$*, condition* (i)*, and Assumption 2.12 are satisfied. Then*

$$LR\left\{ H_\beta(r) \middle| H_{\alpha_\perp^\circ}(r) \right\} \xrightarrow{\mathcal{D}} \chi^2 \left\{ r(p-s) \right\}.$$

Proof of Lemma B.2. If it is argued that

$$S^{-1/2}_{\beta\beta} S_{\alpha\alpha\cdot H,\alpha_\perp} S^{-1/2}_{\beta\beta} \overset{a.s.}{=} S^{-1/2}_{\beta\beta} \Omega_{\alpha\alpha\cdot\alpha_\perp} S^{-1/2}_{\beta\beta} + \mathrm{o}(1). \tag{B.7}$$

$$S_{H_\perp H_\perp \cdot H,\alpha_\perp} - S_{H_\perp H_\perp \cdot H} \overset{a.s.}{=} S_{H_\perp \alpha_\perp \cdot H} S^{-1}_{\alpha_\perp \alpha_\perp \cdot H} S_{\alpha_\perp H_\perp \cdot H} \overset{a.s.}{=} \mathrm{o}(1) \tag{B.8}$$

$$S^{-1/2}_{\beta\beta} S_{\alpha H_\perp \cdot H,\alpha_\perp} S^{-1/2}_{H_\perp H_\perp \cdot H} \overset{a.s.}{=} S^{-1/2}_{\beta\beta} \overline{\alpha}'_\omega S_{\varepsilon H_\perp \cdot H} S^{-1/2}_{H_\perp H_\perp \cdot H} + \mathrm{o}(1). \tag{B.9}$$

then the test statistic can be rewritten as

$$LR\left\{ H_\beta(r) \middle| H_{\alpha_\perp^\circ}(r) \right\} \overset{a.s.}{=} -T \log \det \left(I_p - \Omega^{-1}_{\alpha\alpha\cdot\alpha_\perp} \overline{\alpha}'_\omega S_{\varepsilon H_\perp \cdot H} S^{-1}_{H_\perp H_\perp \cdot H} S_{H_\perp \varepsilon \cdot H} \overline{\alpha}_\omega \right) + \mathrm{o}(1)$$

and the result follows from the Mixed Gaussian Central Limit Theorem A.14.

(B.7): Rewrite $S_{\alpha\alpha\cdot H,\alpha_\perp}$ as

$$S_{\alpha\alpha\cdot H,\alpha_\perp} = S_{\alpha\alpha} - \left(S_{\alpha\alpha_\perp} S^{-1/2}_{\alpha_\perp\alpha_\perp}, S_{\alpha H} S^{-1/2}_{HH} \right) Q^{-1} \begin{pmatrix} S^{-1/2}_{\alpha_\perp\alpha_\perp} S_{\alpha_\perp\alpha} \\ S^{-1/2}_{HH} S_{H\alpha} \end{pmatrix}$$

where

$$Q = \begin{pmatrix} S_{\alpha_\perp\alpha_\perp} & 0 \\ 0 & S_{HH} \end{pmatrix}^{-1/2} \begin{pmatrix} S_{\alpha_\perp\alpha_\perp} & S_{\alpha_\perp H} \\ S_{H\alpha_\perp} & S_{HH} \end{pmatrix} \begin{pmatrix} S_{\alpha_\perp\alpha_\perp} & 0 \\ 0 & S_{HH} \end{pmatrix}^{-1/2}.$$

The equations (A.1), (A.4) of Lemma A.7 show that $Q \overset{a.s.}{=} I_{p-r+s}+\mathrm{o}(1)$ and hence

$$S_{\alpha\alpha\cdot H,\alpha_\perp} \overset{a.s.}{=} \left(S_{\alpha\alpha\cdot\alpha_\perp} - S_{\alpha H} S_{HH}^{-1} S_{H\alpha}\right)\{1+\mathrm{o}(1)\}+\mathrm{o}(1).$$

These equations also imply $S_{\beta\beta}^{-1/2} S_{\alpha H} S_{HH}^{-1} S_{H\alpha} S_{\beta\beta}^{-1/2} \overset{a.s.}{=} I_r+\mathrm{o}(1)$ and $S_{\beta\beta}^{-1/2} S_{\alpha\alpha\cdot\alpha_\perp} S_{\beta\beta}^{-1/2} \overset{a.s.}{=} S_{\beta\beta}^{-1/2}\Omega_{\alpha\alpha\cdot\alpha_\perp} S_{\beta\beta}^{-1/2} + I_r + \mathrm{o}(1)$ and (B.7) follows.

(B.8): This holds since $S_{\alpha_\perp\alpha_\perp\cdot H} \overset{a.s.}{\to} \Omega_{\alpha_\perp\alpha_\perp}$ by (A.1), (A.4) and $S_{\alpha_\perp H_\perp\cdot H} S_{H_\perp H_\perp\cdot H}^{-1/2} = \alpha_\perp' S_{\varepsilon H_\perp\cdot H} S_{H_\perp H_\perp\cdot H}^{-1/2} \overset{a.s.}{\to} 0$ by Lemmas A.5, A.6.

(B.9): The partial model equation (B.6) shows $S_{\alpha H_\perp\cdot H,\alpha_\perp} = \overline{\alpha}_\omega' S_{\varepsilon H_\perp\cdot H,\alpha_\perp}$, hence

$$\begin{aligned} & \overline{\alpha}_\omega'\left(S_{\varepsilon H_\perp\cdot H,\alpha_\perp} - S_{\varepsilon H_\perp\cdot H}\right) S_{H_\perp H_\perp\cdot H}^{-1/2} \\ = \; & \overline{\alpha}_\omega'\left(S_{\varepsilon\varepsilon} - S_{\varepsilon H} S_{HH}^{-1} S_{H\varepsilon}\right)\alpha_\perp\left(S_{\alpha_\perp\alpha_\perp} - S_{\alpha_\perp H} S_{HH}^{-1} S_{H\alpha_\perp}\right)^{-1} S_{\alpha_\perp H_\perp\cdot H} S_{H_\perp H_\perp\cdot H}^{-1/2}, \end{aligned}$$

It has to be argued that the right hand side of this expression converges to zero. First, $\overline{\alpha}_\omega' S_{\varepsilon\alpha_\perp} = \overline{\alpha}_\omega' S_{\varepsilon\varepsilon}\alpha_\perp \overset{a.s.}{=} \mathrm{o}(1)$ by Lemma A.1. Secondly, $S_{\varepsilon H} S_{HH}^{-1/2} \overset{a.s.}{=} \mathrm{o}(1)$ by Lemmas A.5, A.6. Thirdly, $S_{\alpha_\perp\alpha_\perp} \overset{a.s.}{\to} \Omega_{\alpha_\perp\alpha_\perp}$ by (A.1) whereas $S_{\alpha_\perp H} S_{HH}^{-1} S_{H\alpha_\perp} \overset{a.s.}{=} \mathrm{o}(1)$ by (A.4). Fourthly, $S_{\alpha_\perp H_\perp\cdot H} S_{H_\perp H_\perp\cdot H}^{-1/2} \overset{a.s.}{=} \mathrm{o}(1)$ by Lemmas A.5, A.6. Thus (B.9) follows using that $S_{\beta\beta}^{-1/2}$ is finite by Lemma A.4. ■

B.6 *Test for* $\beta = H\varphi$

Proof of Theorem 2.7. Let $L(\alpha_\perp,\beta,\Omega)$ denote the likelihood concentrated with respect to the lagged differences. The likelihood ratio test statistic is then

$$Q\left\{\left.H_\beta(r)\right| H(r)\right\} = \max_{\alpha_\perp,\varphi,\Omega} L(\alpha_\perp, H\varphi,\Omega) \Big/ \max_{\alpha_\perp,\beta,\Omega} L(\alpha_\perp,\beta,\Omega).$$

Extend this fraction by the maximised likelihood when $\alpha_\perp$ is known to see

$$Q\left\{\left.H_\beta(r)\right| H(r)\right\} = \left(\frac{\max_{\alpha_\perp,\varphi,\Omega} L(\alpha_\perp, H\varphi,\Omega)}{\max_{\varphi,\Omega} L(\alpha_\perp^\circ, H\varphi,\Omega)}\right)\left(\frac{\max_{\varphi,\Omega} L(\alpha_\perp^\circ, H\varphi,\Omega)}{\max_{\varphi,\Omega} L(\alpha_\perp^\circ,\beta,\Omega)}\right)\left(\frac{\max_{\varphi,\Omega} L(\alpha_\perp^\circ,\beta,\Omega)}{\max_{\alpha_\perp,\beta,\Omega} L(\alpha_\perp,\beta,\Omega)}\right).$$

Lemma B.1 shows that asymptotically the tests for simple hypotheses on $\alpha_\perp$ only depend on H through β and therefore the first and the last term cancel. The Theorem then follows from Lemma B.2. ■

B.7 *The model with a constant term*

Proof of Theorem 2.9. Because of the full column rank of β a τ_c can be found so $\delta_c = -\tau_c'\beta$. The invariance follows by substituting X_t by X_t° in equation (2.6). ■

Proof of Theorem 2.10. The proof is obtained by minor modifications of the previous proofs. A detailed proof can be obtained from the author.

The estimator $\tilde{\psi}_\perp$ is slightly different in the model $H_\alpha^c(r)$ as compared to the model $H_\alpha(r)$. It involves additional terms of the type $S_{Ac\cdot H} S_{cc\cdot H}^{-1/2}$ which can be dealt with using (A.5) of Lemma A.5. In connection with the expansion for $S_{\beta\beta}^{1/2} U_T$ in (A.9) a term $S_{c\beta} S_{\beta\beta}^{-1/2}$ arises. That term is discussed in Lemma A.10 and the modified version of (A.9) therefore only holds in probability. These results are sufficient to ensure that the Theorems 2.4, 2.5, 2.7, 2.8 can be modified.

By modifying Lemma B.2 the test statistic for $H(r)$ against $H_c(r)$ is found to be

$$LR\left\{H\left(r\right)|H_c\left(r\right)\right\} = \operatorname{tr}\left\{\Omega_{\alpha\alpha\cdot\alpha_\perp}^{-1} T \overline{\alpha}_\omega' S_{\varepsilon c\cdot H} S_{cc\cdot H}^{-1} S_{c\varepsilon\cdot H} \overline{\alpha}_\omega\right\} + \mathrm{o}_{\mathcal{P}}\left(1\right), \tag{B.10}$$

where $R_c = (1|Z_{t-1})$. Lemma A.10 shows that when condition (ii) is satisfied this is

$$LR\left\{H\left(r\right)|H_c\left(r\right)\right\} = \operatorname{tr}\left\{\Omega_{\alpha\alpha\cdot\alpha_\perp}^{-1} T \overline{\alpha}_\omega' S_{\varepsilon c} S_{cc}^{-1} S_{c\varepsilon} \overline{\alpha}_\omega\right\} + \mathrm{o}_{\mathcal{P}}\left(1\right)$$

which is asymptotically χ^2-distributed. When (ii) does not hold a Dickey-Fuller-type distribution arises. ∎

B.8 *The model with a linear term*

Proof of Theorem 2.11. The model equation (2.8) can be re-arranged as

$$\Psi\Delta X_t = \alpha\left(\beta' X_{t-1} + \beta_l' t\right) + \mu_c - \sum_{j=1}^{k} \Gamma_j \sum_{l=1}^{j} \Delta^2 X_{t-l+1} + \varepsilon_t.$$

Replacing X_{t-1} by $X_{t-1}^\circ + \tau_c + \tau_l t$ shows that it has to proved that $\beta'\tau_l + \beta_l' = 0$ and $\mu - \Psi\tau_l + \alpha\beta'\tau_c = 0$. The former follows immediately, whereas insertion of τ_c in the latter shows $\overline{\alpha}_\perp \alpha_\perp' (\mu_c - \Psi\tau_l) = 0$. Replacing τ_l by $(\beta_\perp \overline{\beta}_\perp' + \overline{\beta}\beta')\tau_l$ completes the proof. ∎

C Proofs of Almost Sure Results

Proof of Lemma A.1. It suffices to consider the univariate case. The sum $T^{-1}\sum_{t=1}^{T}(\varepsilon_t^2-\Omega)$ is a martingale and converges to zero almost surely on the set $\{\sum_{t=1}^{\infty}E(\|\varepsilon_t^2-\Omega\|^{1+\gamma/2}|\mathcal{F}_{t-1})<\infty\}$, see Hall and Heyde (1980, Theorem 2.17). This set has probability one by Assumption 2.12. ■

Proof of Lemma A.2. It suffices to consider the univariate case. The sum $T^{\zeta-1}\sum_{t=1}^{T}(\varepsilon_t^2-\Omega)\overset{a.s.}{\to}0$ on the set $\{\sum_{t=1}^{\infty}t^{p(\zeta-1)}E(\|\varepsilon_t^2-\Omega\|^{p}|\mathcal{F}_{t-1})<\infty\}$ for some $1\leq p\leq 2$, see Hall and Heyde (1980, Theorem 2.18). This set has probability one if $p\leq 1+\gamma/2$ and $p(\zeta-1)<-1$ by Assumption 2.12. These three restrictions are satisfied when $\zeta<\min(1/2,\xi)$. The statement for $\sum_{t=1}^{T}\varepsilon_t$ follows likewise. ■

Proof of Lemma A.3. The result follows if for any $\delta>0$ then $t^{-(1-\xi)/2}||\varepsilon_t||>\delta$ only finitely often with probability one, and in particular if $\sum_{t=1}^{\infty}1_{\{||\varepsilon_t||>\delta t^{(1-\xi)/2}\}}\overset{a.s.}{<}\infty$. By the conditional Borel-Cantelli Theorem, see Freedman (1973, Proposition 32), this sum is finite with probability one on the set $[\sum_{t=1}^{\infty}P\{||\varepsilon_t||>t^{(1-\xi)/2}|\mathcal{F}_{t-1}\}<\infty]$. That set is included in $\{\sum_{t=1}^{\infty}t^{-(1-\xi)(1+\gamma/2)}E(||\eta_t||^{2+\gamma}|\mathcal{F}_{t-1})<\infty\}$ by Chebychev's inequality. The latter set has probability one by the martingale difference assumption 2.12. ■

An algebraic lemma is useful.

Lemma C.1 *Consider a sequence of vectors, $\{x_t',y_t'\}'$, $t=1,\ldots,T$. Then*

$$\lambda_{\min}\left\{\sum_{t=1}^{T}(x_t|y_t)^{\otimes 2}\right\}\geq\frac{1}{\dim(x)+\dim(y)}\lambda_{\min}\left\{\sum_{t=1}^{T}\begin{pmatrix}x_t\\ y_t\end{pmatrix}^{\otimes 2}\right\}\geq 0.$$

Proof of Lemma C.1. The inequality holds trivially if the matrix $\sum_{t=1}^{T}\{(x_t',y_t')'\}^{\otimes 2}$ is singular. Therefore assume it is positive definite. In particular the upper left block of its inverse is given by $\{\sum_{t=1}^{T}(x_t|y_t)^{\otimes 2}\}^{-1}$, hence

$$\operatorname{tr}\left[\left\{\sum_{t=1}^{T}(x_t|y_t)^{\otimes 2}\right\}^{-1}\right]\leq\operatorname{tr}\left[\left\{\sum_{t=1}^{T}\begin{pmatrix}x_t\\ y_t\end{pmatrix}^{\otimes 2}\right\}^{-1}\right].$$

Further, for a positive definite matrix A then $\lambda_{\min}(A)=\{\lambda_{\max}(A^{-1})\}^{-1}$ and $\lambda_{\max}(A)\leq\operatorname{tr}(A)\leq\dim(A)\lambda_{\max}(A)$ which leads to the desired inequality. ■

The proof of Lemmas A.4 and A.5 are based on results by Lai and Wei (1982, 1983, 1985). The last of these papers discusses a multivariate stochastic difference equation $\mathbf{x}_t = \mathbf{G}\mathbf{x}_{t-1} + \mathbf{e}_t$ and describes the asymptotic behaviour of $\mathbf{x}_t$ and $\sum_{t=1}^T \mathbf{x}_t\mathbf{x}_t'$. To apply those results define the stacked process $\mathbf{x}_t = (Z_t, X_{t-k}'\beta_\perp, X_{t-k}'\beta)'$,

$$\mathbf{e}_t = \begin{pmatrix} \beta_\perp' \\ \beta' \end{pmatrix} \varepsilon_t, \qquad \mathbf{G} = \begin{pmatrix} I_{p-r} & \beta_\perp'\alpha \\ 0 & I_r + \beta'\alpha \end{pmatrix}, \qquad \text{for } k = 0,$$

and

$$\mathbf{e}_t = \begin{pmatrix} I_p \\ 0 \\ \vdots \\ 0 \end{pmatrix} \varepsilon_t, \qquad \mathbf{G} = \begin{pmatrix} \alpha\beta' + \Gamma_1 & \cdots & \cdots & \alpha\beta' + \Gamma_k & 0 & \alpha \\ I_p & 0 & \cdots & 0 & 0 & 0 \\ 0 & \ddots & \ddots & \vdots & \vdots & \vdots \\ \vdots & & \ddots \; I_p & 0 & 0 & 0 \\ 0 & \cdots & 0 & \beta_\perp' & I_{p-r} & 0 \\ 0 & \cdots & 0 & \beta' & 0 & I_r \end{pmatrix}, \qquad \text{for } k > 0.$$

It has to be argued that the conditions (2.19), (3.11) of Lai and Wei (1985) concerning the covariance structure of $\mathbf{x}_t$ are satisfied.

Lemma C.2 *Suppose Assumption 2.12 and condition* (i) *are satisfied. Then*

$$\liminf_{T\to\infty} \lambda_{\min} \left[E\left\{ \sum_{j=0}^k \mathbf{G}^j \mathbf{e}_{T+j}\mathbf{e}_{T+j}' \left(\mathbf{G}^j\right)' \middle| \mathcal{F}_T \right\} \right] \overset{a.s.}{>} 0,$$

$$\liminf_{T\to\infty} \lambda_{\min} \left\{ \sum_{j=0}^k \mathbf{G}^j \left(\frac{1}{T} \sum_{t=1}^T \mathbf{e}_t\mathbf{e}_t' \right) \left(\mathbf{G}^j\right)' \right\} \overset{a.s.}{>} 0.$$

Proof of Lemma C.2. First the case $k > 0$. By assumption $E(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}) \overset{a.s.}{=} \Omega$ while $T^{-1}\sum_{t=1}^T \varepsilon_t\varepsilon_t' \overset{a.s.}{\to} \Omega$ by Lemma A.1, so it suffices to show $\sum_{j=0}^k \mathbf{G}^j u\Omega u' \left(\mathbf{G}^j\right)' > 0$ where $u = (I_p, 0, \ldots, 0)'$ is an $\{(pk+r) \times r\}$-matrix. It is equivalently to argue that $\text{span}\left(u, \mathbf{G}u, \ldots, \mathbf{G}^k u\right) = \mathbf{R}^{pk+r}$. This holds since

$$\left(u, \mathbf{G}u, \ldots, \mathbf{G}^k u\right) = \begin{pmatrix} M_{11} & M_{12} \\ 0 & (\beta_\perp, \beta)' \end{pmatrix}$$

where M_{11} is a real upper triangular square matrix with ones in the diagonal and since, by assumption, β' has full row rank. For $k = 0$ the same argument can be made with $u = (\beta_\perp, \beta)'$. ■

In the next lemma the asymptotic behaviour of the product moment matrix $\sum_{t=1}^T \{(\mathbf{x}_t', 1)'\}^{\otimes 2}$ is discussed. Following Herstein (1975, p.308) there exists a regular, real matrix M which transforms G into a real, rational canonical form. In particular, M can be chosen so $MGM^{-1} = \text{diag}(D, V)$ is a bivariate block diagonal matrix where the absolute values of the eigenvalues of D and V are respectively larger than one and bounded by one.

Lemma C.3 *Suppose Assumption 2.12 and condition* (i) *are satisfied. Define the vectors* $\tilde{\mathbf{x}}_t = (M\mathbf{x}_t', 1, t/T)'$ *and its non-explosive part* $\tilde{v}_t = (v_t', 1, t/T)'$. *Then*

$$\liminf_{T\to\infty} \lambda_{\min}\left\{\frac{1}{T}\sum_{t=1}^{T}\tilde{\mathbf{x}}_t\tilde{\mathbf{x}}_t'\right\} \overset{a.s.}{>} 0, \tag{C.1}$$

$$\lim_{T\to\infty}\max_{t\leq T}\tilde{v}_t'\left(\sum_{s=1}^{T}\tilde{v}_s\tilde{v}_s'\right)^{-1}\tilde{v}_t \overset{a.s.}{=} 0, \tag{C.2}$$

$$T^{-\eta}\log\lambda_{\max}\left(\sum_{s=1}^{T}\tilde{v}_t\tilde{v}_t'\right) \overset{a.s.}{\to} 0, \qquad \textit{for all } \eta > 0. \tag{C.3}$$

Sketch of Proof of Lemma C.3. Details of the proof can be obtained from the author. It is based on results of Lai and Wei (1985). Note that their assumptions (2.19), (3.11) are satisfied by Lemma C.2.

(C.1): In the proof of Lai and Wei (1985, Theorem 3) the case without a constant term is considered. That proof can be used simply by extending their process with a constant term and noting that $\sum_{t=0}^{T}\varepsilon_t \overset{a.s.}{=} \mathrm{o}(T)$ by a Law of Large Numbers.

(C.2): For the proof of their Lemma 4 Lai and Wei (1985) refer to the proof of Lai and Wei (1983, Theorem 4). In the same way (C.2) can be proved.

(C.3): For simplicity omit the last element of $\tilde{v}_t$. The definition of norms implies

$$\left\|\sum_{t=1}^{T}\begin{pmatrix} v_t \\ 1\end{pmatrix}^{\otimes 2}\right\| \leq \left\|\begin{pmatrix} I_{\dim V} & \sum_{t=1}^{T} v_t \\ 0 & 1\end{pmatrix}^{\otimes 2}\right\|\left\|\begin{Bmatrix} \sum_{t=1}^{T}(v_t|1)^{\otimes 2} & 0 \\ 0 & 1\end{Bmatrix}\right\|.$$

The result then follows by applying Theorem 1 and Corollary 1 of Chan and Wei (1985), respectively, to the two terms on the right hand side. ■

Proof of Lemma A.4. This follows by combining the Lemmas C.3 and C.1. ■

The proof of Lemma A.5 can now be given along the lines of the proof of Lai and Wei (1983, Theorem 1) who consider the case $p = \dim X = 1$.

Proof of Lemma A.5. Define the vectors $\tilde{\mathbf{x}}_t = (d_t', v_t', 1, t)'$ and $\tilde{v}_t = (v_t', 1, t)'$ and the normalisation matrix $N = \mathrm{diag}(D^{2T}, \sum_{t=1}^{T}\tilde{v}_{t-1}^{\otimes 2})^{-1/2}$. It suffices to show

$$T^{(\xi-1)/2}\left(N\sum_{t=1}^{T}\tilde{\mathbf{x}}_{t-1}^{\otimes 2}N'\right)^{-1/2} N\sum_{t=1}^{T}\tilde{\mathbf{x}}_{t-1}\left(\eta_{d,t}', \eta_{v,t}'\right) \overset{a.s.}{\to} 0. \tag{C.4}$$

It is first argued that $N\sum_{t=1}^{T}\tilde{\mathbf{x}}_{t-1}^{\otimes 2}N' \overset{a.s.}{\to} \mathrm{diag}(F_D, I)$ where $F_D \overset{a.s.}{>} 0$. The convergence of the top left corner follows from Lai and Wei (1985, Corollary 2). The lower right corner is an identity. For the off diagonal elements use that

$$\left\| D^{-T} \sum_{t=1}^{T} d_{t-1} \tilde{v}_{t-1}' \left(\sum_{s=1}^{T} \tilde{v}_{s-1}^{\otimes 2} \right)^{-1/2} \right\| \leq \sum_{t=1}^{T} \left\| D^{-T} d_{t-1} \right\| \left\| \left(\sum_{s=1}^{T} \tilde{v}_{s-1}^{\otimes 2} \right)^{-1/2} \tilde{v}_{t-1} \right\|$$

This converges to zero by (C.2) of Lemma C.3 and since

$$\sum_{t=1}^{T} \left\| D^{-T} d_{t-1} \right\| \overset{a.s.}{<} \infty \tag{C.5}$$

by Lemma 4 and Corollary 4 of Lai and Wei (1985) using Assumption 2.12.

The second of the matrices in (C.4) has two elements. The first is

$$\left\| D^{-T} \sum_{t=1}^{T} d_{t-1} \eta_t \right\| \leq \left(\max_{t \leq T} \| \eta_t \| \right) \sum_{t=1}^{T} \left\| D^{-T} d_{t-1} \right\| \overset{a.s.}{=} \mathrm{o}\left(T^{1-\xi/2} \right) \mathrm{O}\left(1\right),$$

by Lemma A.3 and (C.5). For the second it suffices to prove that

$$\sum_{t=1}^{T} \eta_t \tilde{v}_{t-1} \left(\sum_{s=1}^{T} \tilde{v}_{s-1}^{\otimes 2} \right)^{-1} \sum_{t=1}^{T} \tilde{v}_{t-1} \eta_t' \overset{a.s.}{=} \mathrm{o}\left(T^{\eta} \right). \tag{C.6}$$

Using (C.3) of Lemma C.3 the result (C.6) then follows from Lemma 1 of Lai and Wei (1982). Their result is actually stated for the case where η_t is one-dimensional but is easily generalised to the multivariate case. ■

Proof of Lemma A.6. Define the matrix

$$M = \left\{ \begin{array}{cc} I_{\dim x} & -\sum_{t=1}^{T} x_t y_t \left(\sum_{t=1}^{T} y_t^{\otimes 2} \right)^{-1} \\ 0 & I_{\dim y} \end{array} \right\}.$$

The result then follows from the identity

$$\left\{ \sum_{t=1}^{T} \left(\begin{array}{c} x_t \\ y_t \end{array} \right)^{\otimes 2} \right\}^{-1/2} \sum_{t=1}^{T} \left(\begin{array}{c} x_t \\ y_t \end{array} \right) a_t' = \left\{ \sum_{t=1}^{T} M \left(\begin{array}{c} x_t \\ y_t \end{array} \right)^{\otimes 2} M' \right\}^{-1/2} M \sum_{t=1}^{T} \left(\begin{array}{c} x_t \\ y_t \end{array} \right) a_t'$$

noting that the latter inverse matrix is block diagonal. ■

Proof of Lemma A.7. (A.1): The concentrated model equation $R_{0,t} = \alpha\beta' R_{1,t} + R_{\varepsilon,t}$ shows that $S_{00} = S_{\varepsilon\varepsilon} + \alpha S_{\beta\varepsilon} + S_{\varepsilon\beta}\alpha' + \alpha S_{\beta\beta}\alpha'$. Pre- and post-multiply this equation by A_T. The first term equals $A_T'\Omega A_T + \mathrm{o}(1)$ by Lemmas A.1, A.5, A.6.

For the second and third term use that $||A_T|| \overset{a.s.}{<} \infty$ by Lemma A.4 and $S_{\varepsilon\beta}\alpha' A_T = (0, S_{\varepsilon\beta}S_{\beta\beta}^{-1/2}) \overset{a.s.}{=} \{0, \mathrm{o}(T^{-\xi/2})\}$ by Lemmas A.5, A.6. Finally, $A_T'\alpha S_{\beta\beta}\alpha' A_T = \mathrm{diag}(0, I_r)$.

(A.2), (A.3): The concentrated model equation shows $\beta' S_{10} = S_{\beta\beta}\alpha' + S_{\beta\varepsilon}$ so $(0, I_r)B_T' S_{10} A_T = (0, I_r) + S_{\beta\beta}^{-1/2} S_{\beta\varepsilon} A_T$ and the result follows as above.

(A.4): Noting that $\mathrm{span}(H) = \mathrm{span}(H\varphi, H\varphi_\perp)$ and $\beta = H\varphi$ it follows that

$$S_{0H} S_{HH}^{-1} S_{H0} = S_{0H\cdot\beta}\varphi_\perp \left(\varphi_\perp' S_{HH\cdot\beta}\varphi_\perp\right)^{-1} \varphi_\perp' S_{H0\cdot\beta} + S_{0\beta} S_{\beta\beta}^{-1} S_{\beta 0}.$$

By the concentrated model equation this equals

$$S_{\varepsilon H\cdot\beta}\varphi_\perp \left(\varphi_\perp' S_{HH\cdot\beta}\varphi_\perp\right)^{-1} \varphi_\perp' S_{H\varepsilon\cdot\beta} + \left(\alpha + S_{\varepsilon\beta} S_{\beta\beta}^{-1}\right) S_{\beta\beta} \left(S_{\beta\beta}^{-1} S_{\beta\varepsilon} + \alpha'\right).$$

Pre- and post-multiply this by A_T and use Lemmas A.4, A.5, A.6. The first term is of order $\mathrm{o}(T^{-\xi})$, whereas for the second term $A_T'\alpha S_{\beta\beta}\alpha' A_T = \mathrm{diag}(0, I_r)$, $A_T'\alpha S_{\beta\varepsilon} A_T \overset{a.s.}{=} \mathrm{o}(T^{-\xi/2})$, $A_T' S_{\varepsilon\beta} S_{\beta\beta}^{-1} S_{\beta\varepsilon} A_T \overset{a.s.}{=} \mathrm{o}(T^{-\xi})$.

(A.5): Again the concentrated model equation shows

$$\sum_{t=1}^{T} \Delta X_t \left(1 | X_{t-1}, Z_{t-1}\right) = \sum_{t=1}^{T} \varepsilon_t \left(1 | X_{t-1}, Z_{t-1}\right)$$

and the formula follows from Lemmas A.5, A.6. ■

Proof of Corollary A.8.

The inversion formula for (2×2)-block matrices implies that

$$\tilde{S}_{00}^{-1} - \tilde{S}_{00}^{-1} \begin{pmatrix} 0 \\ I_r \end{pmatrix} \left\{ (0, I_r)\, \tilde{S}_{00}^{-1} \begin{pmatrix} 0 \\ I_r \end{pmatrix} \right\}^{-1} (0, I_r)\, \tilde{S}_{00}^{-1}$$
$$= \begin{pmatrix} I_{p-r} \\ 0 \end{pmatrix} \left\{ (I_{p-r}, 0)\, \tilde{S}_{00} \begin{pmatrix} I_{p-r} \\ 0 \end{pmatrix} \right\}^{-1} (I_{p-r}, 0). \qquad \text{(C.7)}$$

It has to be shown that N and the left hand side of (C.7) have the same limit. Lemma A.7 shows that $(0, I_r)\, \tilde{S}_{10} \overset{a.s.}{\to} (0, I_r)$. It therefore suffices to argue that the inverse matrices occurring on the left hand side of (C.7) are bounded with probability one. The Lemmas A.4, A.7 show that $\tilde{S}_{00}$ is positive definite in the limit while

$$\left\{ (0, I_r)\, \tilde{S}_{00}^{-1} \begin{pmatrix} 0 \\ I_r \end{pmatrix} \right\}^{-1} \overset{a.s.}{=} I_r + S_{\beta\beta}^{-1/2} \Omega_{\alpha\alpha\cdot\alpha_\perp} S_{\beta\beta}^{-1/2} + \mathrm{o}(1) \overset{a.s.}{=} I_r + \mathrm{O}(1) + \mathrm{o}(1).$$

■

Proof of Lemma A.9. It is convenient to define $M_\psi = (\psi_\perp, \overline{\psi} S_{\beta\beta}^{-1/2})$. Equation (A.1) of Lemma A.5 can be strengthened using Lemma A.2 instead of Lemma A.1,

$$M_\psi' S_{AA\cdot A_\perp} M_\psi \overset{a.s.}{=} M_\psi' \Omega_{AA\cdot A_\perp} M_\psi + \begin{pmatrix} 0 & 0 \\ 0 & I_r \end{pmatrix} + \mathrm{o}\left(T^{-\xi/2}\right). \qquad \text{(C.8)}$$

This formula together with (A.4) of Lemma A.7 imply

$$S_{HH\cdot A_\perp} \overset{a.s.}{=} S_{HH}\left\{1+\mathrm{o}\left(T^{-\xi}\right)\right\}, \tag{C.9}$$

$$M_\psi' S_{AH\cdot A_\perp} S_{HH\cdot A_\perp}^{-1} S_{HA\cdot A_\perp} M_\psi \overset{a.s.}{=} \begin{pmatrix} 0 & 0 \\ 0 & I_r \end{pmatrix} + \left\{\begin{array}{cc} \mathrm{o}\left(T^{-\xi}\right) & \mathrm{o}\left(T^{-\xi/2}\right) \\ \mathrm{o}\left(T^{-\xi/2}\right) & \mathrm{o}\left(T^{-\xi/2}\right) \end{array}\right\} \tag{C.10}$$

(A.7): The concentrated model equation (2.1) is

$$R_{0,t} = A\psi\varphi' H' R_{1,t} + \hat{\varepsilon}_t. \tag{C.11}$$

Pre-multiplication by $(\overline{A}, A_\perp')$ shows that $A_\perp' X_t$ is weakly exogenous for ψ, φ. Thus likelihood inference can be based on the partial system of $\overline{A}' X_t$ given $A_\perp' X_t$, that is

$$\overline{A}' R_{0,t} = \psi\varphi' H' R_{1,t} + \omega A_\perp' R_{0,t} + \overline{A}_\omega' \hat{\varepsilon}_t. \tag{C.12}$$

for $\overline{A}_\omega' = (\overline{A}' - \omega A_\perp')$ and $\omega = \Omega_{AA_\perp}\Omega_{A_\perp A_\perp}^{-1}$. The squared sample canonical correlations for $\overline{A}' R_{0,t}$ and $H' R_{1,t}$ solve $0 = \det M_\psi'(\lambda S_{AA\cdot A_\perp} - S_{AH\cdot A_\perp} S_{HH\cdot A_\perp}^{-1} S_{HA\cdot A_\perp}) M_\psi$ or equivalently $0 \overset{a.s.}{=} \det\{P_\psi(\lambda) + \mathrm{o}(T^{-\xi/2})\}$ where (C.8), (C.10) have been used and $P_\psi(\lambda) = \lambda M_\psi' \Omega_{AA\cdot A_\perp} M_\psi + (\lambda - 1)\mathrm{diag}(0, I_r)$. For $\lambda = 0$ then $(I_{m-r}, 0) P_A(\lambda) = 0$ which shows that

$$\mathrm{span}\left\{\begin{pmatrix} \overline{\psi}_\perp' \\ S_{\beta\beta}^{1/2}\psi' \end{pmatrix} \hat{\psi}_\perp\right\} \overset{a.s.}{=} \mathrm{span}\begin{pmatrix} I_{m-r} \\ 0 \end{pmatrix} + \mathrm{o}\left(T^{-\xi/2}\right). \tag{C.13}$$

Here $\hat{\psi}_\perp$ can be replaced $\tilde{\psi}_\perp$ since $\overline{\psi}_\perp' \hat{\psi}_\perp$ has full rank almost surely.

(A.8): By (A.6) it holds $\tilde{\psi}_\perp = \psi_\perp + \overline{\psi} S_{\beta\beta}^{-1/2} S_{\beta\beta}^{1/2} \overline{\psi}_\perp' \hat{\psi}_\perp$ and the result follows by the boundedness of $S_{\beta\beta}^{-1/2}$, see Lemma A.4.

(A.9): The partial likelihood given by (C.12) is differentiated with respect to ψ in the direction a following Johansen (1996, (13.9)) to obtain the likelihood equation

$$\frac{\partial}{\partial\psi} \log L(\psi, \varphi, \Omega) = T\mathrm{tr}\left\{\Omega_{AA\cdot A_\perp}^{-1}\left(S_{AH\cdot A_\perp} - \psi\varphi' S_{HH\cdot A_\perp}\right)\varphi a'\right\} = 0.$$

In particular it holds that $S_{AH\cdot A_\perp}\hat{\varphi} = \hat{\psi}\hat{\varphi}' S_{HH\cdot A_\perp}\hat{\varphi}$. Since $\tilde{\psi}_\perp'\hat{\psi} = 0$ and ψ has full column rank by condition (i) then $0 = \tilde{\psi}_\perp' S_{AH\cdot A_\perp}\hat{\varphi}\hat{\psi}'\overline{\psi}$. Following Johansen (1996, p. 130) the maximum likelihood estimator for φ is $\hat{\varphi} = S_{HH\cdot\hat{\alpha}_\perp}^{-1} S_{HA\cdot\hat{\alpha}_\perp}\hat{\psi}\left(\hat{\psi}'\hat{\psi}\right)^{-1}$ and therefore $0 = \tilde{\psi}_\perp' S_{AH\cdot A_\perp} S_{HH\cdot\hat{\alpha}_\perp}^{-1} S_{HA\cdot\hat{\alpha}_\perp}\hat{\psi}(\hat{\psi}'\hat{\psi})^{-1}\hat{\psi}\overline{\psi}$. The consistency of $\tilde{\psi}_\perp$ given in (A.7) implies $\hat{\psi}(\hat{\psi}'\hat{\psi})^{-1}\hat{\psi}\overline{\psi} \overset{a.s.}{=} \overline{\psi}\{1+\mathrm{o}(T^{-\xi/2})\}$. Using the concentrated model equation (C.11) it then follows that

$$0 \overset{a.s.}{=} \tilde{\psi}_\perp'\left(\psi\varphi' S_{HH\cdot A_\perp} + \overline{A}' S_{\varepsilon H\cdot A_\perp}\right)\left(\varphi\psi' + S_{HH\cdot\hat{\alpha}_\perp}^{-1} S_{H\varepsilon\cdot\hat{\alpha}_\perp}\overline{A}\right)\overline{\psi}\left\{1+\mathrm{o}\left(T^{-\xi/2}\right)\right\}. \tag{C.14}$$

A few convergence results are needed in order to rewrite (C.14) further. It follows from (A.7), (A.8) that $\tilde{\psi}_{\perp}' M_{\psi}^{-1} = \tilde{\psi}_{\perp}'(\overline{\psi}_{\perp}, \psi S_{\beta\beta}^{1/2}) \overset{a.s.}{=} (I_{m-r}, 0) + \mathrm{o}(T^{-\xi/2})$. This in combination with Lemma A.4, (C.8) and (C.10) shows that

$$\tilde{\psi}_{\perp}' S_{AA\cdot A_{\perp}} \tilde{\psi}_{\perp} \overset{a.s.}{=} \psi_{\perp}' \Omega_{AA\cdot A_{\perp}} \psi_{\perp} + \mathrm{o}(T^{-\xi/2}), \tag{C.15}$$

$$\tilde{\psi}_{\perp}' S_{AH\cdot A_{\perp}} S_{HH\cdot A_{\perp}}^{-1} S_{HA\cdot A_{\perp}} \tilde{\psi}_{\perp} \overset{a.s.}{=} \mathrm{o}\left(T^{-\xi}\right), \tag{C.16}$$

while the concentrated model equation (C.11) and Lemmas A.1, A.5, A.6 give

$$\tilde{\psi}_{\perp}' S_{A\varepsilon\cdot A_{\perp}} \overset{a.s.}{=} \mathrm{O}\left(1\right). \tag{C.17}$$

Using (C.15)-(C.17) it is seen that $S_{HH\cdot\hat{\alpha}} \overset{a.s.}{=} S_{HH\cdot A_{\perp}}\{1+\mathrm{o}(T^{-\xi})\}$ and $S_{HH\cdot A_{\perp}}^{-1/2} S_{H\varepsilon\cdot\hat{\alpha}_{\perp}} \overset{a.s.}{=} S_{HH\cdot A_{\perp}}^{-1/2} S_{H\varepsilon\cdot A_{\perp}} + \mathrm{o}(T^{-\xi/2})$. It follows that (C.14) is equivalent to

$$\begin{aligned} 0 \overset{a.s.}{=} & \tilde{\psi}_{\perp}' \left[\left(\psi\varphi' S_{HH\cdot A_{\perp}} + \overline{A}' S_{\varepsilon H\cdot A_{\perp}}\right)\left\{S_{HH\cdot A_{\perp}}^{-1} S_{H\varepsilon\cdot A_{\perp}} \overline{A\psi} + S_{HH\cdot A_{\perp}}^{-1/2} \mathrm{o}\left(T^{-\xi/2}\right)\right\}\right. \\ & \left. + \psi S_{\beta\beta\cdot A_{\perp}} + \overline{A}' S_{\varepsilon\beta\cdot A_{\perp}}\right]\left\{1 + \mathrm{o}\left(T^{-\xi/2}\right)\right\} \end{aligned}$$

The Lemmas A.1, A.5, A.6 show $S_{HH}^{-1/2} S_{H\varepsilon\cdot A_{\perp}} = \mathrm{o}(T^{-\xi/2})$. Therefore using (A.8), (C.8) and (A.1) of Lemma A.7

$$\tilde{\psi}_{\perp}' \psi S_{\beta\beta} \overset{a.s.}{=} -\psi_{\perp}' \overline{A}_{\omega}' S_{\varepsilon\beta}\left\{1 + \mathrm{o}\left(T^{-\xi/2}\right)\right\} + \mathrm{o}(T^{-\xi}). \tag{C.18}$$

Finally, use that $S_{\beta\beta}^{-1/2}$ is bounded by Lemma A.4. ■

D Proof of Convergence in Probability Result

As a first step towards the proof of Lemma A.10 define the process $\mathbf{u}_t = (X_t'\beta, Z_t')'$. Following the argument preceding Lemma C.3 a real, regular $(pk+p)$-matrix M can be found so $M\mathbf{u}_t = (a_t', b_{1t}', \ldots, b_{lt}', c_t', d_t')'$, $M\mathbf{G}M^{-1} = \mathrm{diag}(A, B_1, \ldots, B_l, C, D)$, and $M\mathbf{e}_t = (\eta_{at}', \eta_{b_1 t}', \ldots, \eta_{b_l t}', \eta_{ct}', \eta_{dt}')'$, where A has roots only at one, B_j has roots on the unit circle at the complex pairs $\exp(i\theta_j)$ and $\exp(-i\theta_j)$ for some $0 < \theta_j \leq \pi$ while C and D have roots with modulus smaller and larger than one, respectively. Moreover, the innovations $M\mathbf{e}_t$ satisfy the property described in Lemma C.2.

Two Lemmas are then needed.

Lemma D.1 *Suppose equation* (2.1)*, the hypothesis* $H(r)$*, condition* (i)*, and Assumption 2.12 are satisfied. Then there exist normalisations* N_a, N_{b_j}, N_c, N_d *so*

$$\sum_{t=1}^{T} \{\mathrm{diag}\,(N_a, N_{b_1}, \ldots, N_{b_l}, N_c, N_d)\, M\mathbf{u}_t\}^{\otimes 2} \xrightarrow{\mathcal{D}} F = \mathrm{diag}\,(F_a, F_{b_1}, \ldots, F_{b_l}, F_c, F_d)\,.$$

The limiting matrix F *is possibly random, but positive definite with probability one.*

Sketch of Proof of Lemma D.1. This Lemma generalises results by Chan and Wei (1988, Section 3). Details can be obtained from the author.

First the convergence $\sum_{t=1}^{T}\{N_a a_t\}^{\otimes 2} \xrightarrow{\mathcal{D}} F_a$ is considered. A real, regular similarity transformation can be chosen so that A is transformed into a Jordan form, that is a block diagonal matrix with Jordan blocks

$$\begin{pmatrix} 1 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 1 \end{pmatrix},$$

see Herstein (1975, p. 302, 312). For simplicity suppose A is a Jordan block. As in Lemma C.2 it follows that $E\{\sum_{j=0}^{k} A^j \eta_{a,T+j}\eta_{a,T+j}'(A')^j | \mathcal{F}_T\} \overset{a.s.}{>} 0$, implying that the conditional variance of the last element of η_a is positive. A suitable normalisation matrix N_a and a limit covariance matrix F_a can then be found along the lines of Chan and Wei (1988, Section 3.1).

For the component b_j a corresponding argument is made. When the roots are complex the Jordan matrix is complex so it is convenient to work with blocks of the form

$$\begin{pmatrix} B & I_2 & & \\ & \ddots & \ddots & \\ & & \ddots & I_2 \\ & & & B \end{pmatrix} \qquad \text{where} \qquad B = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

The cases where the roots are either smaller or larger than one are dealt with by Lai and Wei (1985, Theorem 2 and Equation 3.9) and Lai and Wei (1985, Corollary 2) respectively. See also the proof of Lemma A.5.

The cross product terms involving the explosive roots are dealt with in the proof of Lemma A.5. For the remaining cross product terms combine Chan and Wei (1988, Theorem 3.4.1, 3.4.2) with the Jordan type decompositions used above. ■

Lemma D.2 *Suppose equation* (2.1)*, the hypothesis,* $H(r)$*, condition* (i)*, and Assumption 2.12 are satisfied. Let* y_t *be either* $(t/T)^k$ *or* $T^{-k-1/2}\sum_{s=1}^{t-1}(t-s)^k\varepsilon_s$ *for* $k=0,1,\ldots$ *If* u_t *is given as either of the processes* $b_{j,t},c_t,d_t$ *then for all* $\eta>0$

$$T^{-\eta}\left(\sum_{t=1}^{T}u_{t-1}u_{t-1}'\right)^{-1/2}\sum_{t=1}^{T}u_{t-1}y_t'\overset{\mathcal{P}}{\to}0$$

whereas

$$T^{-1/2}\left(\sum_{t=1}^{T}a_{t-1}a_{t-1}'\right)^{-1/2}\sum_{t=1}^{T}N_a a_{t-1}y_t'=\mathrm{O}_{\mathcal{P}}\left(1\right).$$

Sketch of Proof of Lemma D.2. For the first result the argument is essentially the same as in the proof of Nielsen (2000, Lemmas A.4 and B.1) and is based on results and ideas of Chan and Wei (1988). The second result follows using the Continuous Mapping Theorem since $T^{1/2}N_a a_{[Tu]}$ and $y_{[Tu]}$ converge jointly in distribution. Details can be obtained from the author. ■

Proof of Lemma A.10. Condition (ii) implies that the process $\mathbf{u}_t=(X_t'\beta,Z_t')'$ has no unit roots. This is because the companion matrix for $\tilde{\mathbf{u}}_t=(X_t'\beta_\perp,X_t'\beta,Z_t')'$ is upper triangular, showing that $(p-r)$ unit roots are associated with $\beta_\perp'X_t$ and at the same time $\tilde{\mathbf{u}}_t$ has at most $(p-r)$ unit roots as argued in Johansen (1996, Section 4). The proof now follows from the Lemmas D.1, D.2 noting that

$$\left(\sum_{t=1}^{T}u_{t-1}u_{t-1}'\right)^{-1/2}\sum_{t=1}^{T}u_{t-1}y_t'=\left(M'\sum_{t=1}^{T}u_{t-1}u_{t-1}'M\right)^{-1/2}M'\sum_{t=1}^{T}u_{t-1}y_t'$$

■

E Proofs of Distributional Results

Proof of Lemma A.11. See Chan and Wei (1988, Theorem 2.2). ■

For the proof of Lemma A.12 it is convenient to introduce the notation $S_t = \sum_{s=0}^{t} \varepsilon_s$ where $\varepsilon_0 = \alpha'_\perp \Psi X_0 + \sum_{j=1}^{k} \sum_{l=j}^{k} \alpha'_\perp \Gamma_l \Delta X_{1-j}$.

Lemma E.1 *Suppose equation* (2.1)*, the hypothesis* $H(r)$ *and the conditions* (i), (ii) *are satisfied. Then* $\alpha'_\perp \Psi R_{1,t} = \alpha'_\perp (S_t | Z_{t-1})$.

Proof of Lemma E.1. Without loss of generality let $k = 1$. The model equation (2.1) is equivalent to $\Psi \Delta X_t = \alpha\beta' X_{t-1} - \Gamma_1 \Delta^2 X_{t-1} + \varepsilon_t$. Pre-multiplication with $\alpha'_\perp$ and summation gives $\alpha'_\perp \Psi X_{t-1} = \alpha'_\perp S_{t-1} - \alpha'_\perp \Gamma_1 \Delta X_{t-1}$ and the result follows. ■

Proof of Theorem A.12. Using the Lemma E.1 and the model equation (2.1) it follows that $(I_{p-r}, 0) C'_T S_{10} = T^{-1} \sum_{t=1}^{T} (\alpha'_\perp S_{t-1} | \beta' X_{t-1}, Z_{t-1}) \varepsilon'_t$. With the choice $0 < \eta < \xi < \gamma/(2+\gamma) < 1$ the Lemmas A.5, A.10 imply $(I_{p-r}, 0) C'_T S_{10} A_T M_T = T^{-1} \sum_{t=1}^{T} \alpha'_\perp S_{t-1} \varepsilon'_t + o_{\mathcal{P}}(1)$. In the same way, $C'_T S_{11} C_T = \operatorname{diag}(T^{-1} \sum_{t=1}^{T} \alpha'_\perp S_{t-1}^{\otimes 2} \alpha_\perp, I_r)$ $+ o_{\mathcal{P}}(1)$. The result then follows from Lemma A.11. ■

Proof of Lemma A.13. The result follows from Brown (1971, Theorem 2, Lemma 2). It suffices to consider the univariate case. Define $x_t = \varepsilon_t \varepsilon_{t-1}$ and note that the second moment of x_t is finite since the fourth moment of ε_t is finite. Let $S_t = \sum_{r=1}^{t} x_r^2$, $\sigma_t^2 = E(x_t^2 | \mathcal{F}_{t-1}) = \Omega \varepsilon_{t-1}^2$, $V_t^2 = \sum_{r=1}^{t} \sigma_r^2 = \Omega \sum_{r=1}^{t} \varepsilon_{r-1}^2$, and $s_t^2 = EV_t^2 = t\Omega^2$. By the Law of Large Numbers given in Lemma A.1 it follows that $V_t^2 / s_t^2 \xrightarrow{\mathcal{P}} 1$. Further a conditional Lindeberg condition has to be satisfied. To this end note

$$\frac{1}{V_t^2} \sum_{r=1}^{t} E\left\{ x_r^2 1_{\left(x_r^2 \geq \delta s_t^2\right)} \middle| \mathcal{F}_{r-1} \right\} = \frac{1}{V_t^2} \sum_{r=1}^{t} \varepsilon_{r-1}^2 E\left\{ \varepsilon_r^2 1_{\left(\varepsilon_r^2 \varepsilon_{r-1}^2 \geq \delta s_t^2\right)} \middle| \mathcal{F}_{r-1} \right\}. \tag{E.1}$$

By a Chebychev inequality type argument this is bounded for all $0 \leq \eta \leq \gamma$ by $(V_t^2 \delta^\eta s_t^\eta)^{-1} \sum_{r=1}^{t} ||\varepsilon_{r-1}||^{2+\eta} E\{||\varepsilon_r||^{2+\eta} | \mathcal{F}_{r-1}\}$. By Assumption 2.12 this is bounded by some constant times $(V_t^2 t^\eta)^{-1} \sum_{r=1}^{t} ||\varepsilon_{r-1}||^{2+\eta}$. As for the denominator of this expression note that $t^{-1} V_t^2 \xrightarrow{\mathcal{P}} \Omega^2$ by the Law of Large Numbers given in Lemma A.1. The numerator can be rewritten as the sum of the martingale $t^{-(1+\eta)} \sum_{r=1}^{t} \{||\varepsilon_r||^{2+\eta} - E(||\varepsilon_r||^{2+\eta} | \mathcal{F}_{r-1})\}$ and $t^{-(1+\eta)} \sum_{r=1}^{t} E(||\varepsilon_r||^{2+\eta} | \mathcal{F}_{r-1})$. The first term converges to zero by the same argument as in the proof of Lemma A.1 whereas the second converges to zero by Assumption 2.12. This shows that the expression (E.1) converges to zero almost surely. Brown's result then implies that $S_t / s_t = (\sum_{r=1}^{t} \varepsilon_t \varepsilon_{t-1}) / (t^{1/2} \Omega) \xrightarrow{\mathcal{D}} N(0, 1)$. ■

Proof of Lemma A.14. First, consider the case where $H = \beta$, $H_\perp = \beta_\perp$. Theorem 2.3 is formulated in terms of $\beta_1, \ldots, \beta_m$. Together with Lemma A.10 and Cauchy-Schwartz's inequality it implies that, for $1 \leq j, l \leq m$,

$$\begin{aligned} \beta_j' S_{11} \beta_l &= \sum_{t=1}^{T} \left(\frac{\alpha_j}{j!} \sum_{s=1}^{t-1} s^{j-1} \varepsilon_s \middle| \beta' X_{t-1}, Z_{t-1} \right) \left(\frac{\alpha_l}{l!} \sum_{s=1}^{t-1} s^{l-1} \varepsilon_s \right)' + \mathrm{o}_{\mathcal{P}} \left(T^{j+l} \right), \\ &= \sum_{t=1}^{T} \left(\frac{\alpha_j}{j!} \sum_{s=1}^{t-1} s^{j-1} \varepsilon_s \right) \left(\frac{\alpha_l}{l!} \sum_{s=1}^{t-1} s^{l-1} \varepsilon_s \right)' + \mathrm{o}_{\mathcal{P}} \left(T^{j+l} \right). \end{aligned}$$

Correspondingly, using Lemmas A.5, A.6 it also holds that

$$\beta_j' S_{1\varepsilon \cdot \beta} = \sum_{t=1}^{T} \left(\frac{\alpha_j}{j!} \sum_{s=1}^{t-1} s^{j-1} \varepsilon_s \right) \varepsilon_t' + \mathrm{o}_p \left(T^j \right).$$

Lemma A.11 in conjunction with the Continuous Mapping Theorem, see Billingsley (1968) then shows

$$\Omega_{\alpha\alpha\cdot\alpha_\perp}^{-1/2} \overline{\alpha}_\omega' S_{\varepsilon 1 \cdot \beta} S_{11\cdot\beta}^{-1/2} \xrightarrow{\mathcal{D}} \Omega_{\alpha\alpha\cdot\alpha_\perp}^{-1/2} \overline{\alpha}_\omega' \int_0^1 dW_u g\left(\alpha_\perp' W_u\right) \left[\int_0^1 \left\{ g\left(\alpha_\perp' W_u\right) g\left(\alpha_\perp' W_u\right)' \right\} du \right]^{-1/2}$$

where W is a p-dimensional Brownian motion with variance Ω and g is some continuous function. The Brownian motions $\overline{\alpha}_\omega' W$ and $\alpha_\perp' W$ are independent. Therefore by conditioning on $\alpha_\perp' W$ this variable is seen to have a $\{r(p-r)\}$-dimensional standard normal distribution.

In the general case where $\beta = H\varphi$ let $\beta_\perp = (H_\perp, H\varphi_\perp)$. Then $\left(H_\perp' R_{1,t} \middle| H' R_{1,t} \right) = \left\{ I_{p-r}, -S_{H_\perp H\cdot\beta} \varphi_\perp \left(\varphi_\perp' S_{HH\cdot\beta} \varphi_\perp \right)^{-1} \right\} \left(\beta_\perp' R_{1,t} \middle| \beta' R_{1,t} \right)$ and the above approach can be used. ■

References

Anderson, T.W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Annals of Mathematical Statistics* 22, 327-351. Correction in *Annals of Statistics* 8, 1400 (1980).

Bartlett, M.S. (1938). Further aspects of the theory of multiple regression. *Proceedings of the Cambridge Philosophical Society* 34, 33-40.

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

Brown, B.M. (1971). Martingale Central Limit Theorems. *Annals of Mathematical Statistics* 42, 59-66.

Chan, N.H., and C.Z. Wei (1988). Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes. *Annals of Statistics* 16, 367-401.

la Cour, L. (1998). A parametric characterization of integrated vector autoregressive (VAR) processes. *Econometric Theory 14, 187-199.*

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987-1007.

Engle, R.F. and Granger, C.W.J. (1987). Co-integration and error correction: Representation, estimation and testing. *Econometrica* 55, 251-276.

Freedman, D. (1973). Another note on the Borel-Cantelli Lemma and the strong law, with the Poisson approximation as a by-product. *Annals of Probability* 1, 910-925.

Hall, P., and Heyde, C.C. (1980). *Martingale Limit Theory and Its Applications.* Academic Press, San Diego.

Hansen, B.E. (1992). Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory* 8, 489-500

Herstein, I.N. (1975). *Topics in Algebra*, 2nd edition. Wiley, New York.

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

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

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

Johansen, S. (1996). *Likelihood-based inference in cointegrated vector autoregressive models.* 2nd printing. Oxford University Press.

Juselius, K. and Mladenovic, Z. (1999). High inflation, hyper inflation and explosive roots. The case of Jugoslavia. Mimeo.

Lai, T.L. and Wei, C.Z. (1982). Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Annals of Statistics* 10, 154-166.

Lai, T.L. and Wei, C.Z. (1983). Asymptotic Properties of General Autoregressive Models and Strong Consistency of Least-Squares Estimates of Their Parameters. *Journal of Multivariate Analysis* 13, 1-23.

Lai, T.L. and Wei, C.Z. (1985). Asymptotic properties of multivariate weighted sums with applications to stochastic regression in linear dynamic systems. In P.R. Krishnaiah, ed., *Multivariate Analysis VI*, Elsevier Science Publishers, 375-393.

Lee, H.S. (1992). Maximum likelihood inference in cointegration and seasonal cointegration. *Journal of Econometrics* 54, 1-47.

Nielsen, B. (2000). The asymptotic distribution of unit root tests of unstable autoregressive processes. To appear in *Econometrica.*

Nielsen, B. and Rahbek, A. (2000). Similarity issues in cointegration analysis. *Oxford Bulletin of Economics and Statistics* 62, 5-22.

Paroulo, P. (1994). The role of the drift in I(2) systems. *Journal of the Italian Statistical Society* 3, 93-123. Correction by Paroulo, P. and Nielsen, B. (1999) manuscript.

Phillips, P.C.B. and Solo, V. (1992). Asymptotics for linear processes. *Annals of Statistics* 20, 971-1001.

Rahbek, A., Kongsted, H.C., and Jørgensen, C. (1999). Trend stationarity in the I(2) cointegration model. *Journal of Econometrics* 90, 265-289.

THE ASYMPTOTIC DISTRIBUTION OF THE LIKELIHOOD RATIO TEST STATISTICS FOR COINTEGRATION IN UNSTABLE VECTOR AUTOREGRESSIVE PROCESSES

DETAILS OF PROOFS

By Bent Nielsen

4 September 2000

Index:

F.1 *Order of magnitude of* $\sum_{t=1}^{T} \mathbf{x}_{t-1}\mathbf{x}_{t-1}'$

Lemma F.1 (Lemma C.3) *Suppose the martingale difference assumption 2.12 and condition* (i) *are satisfied. Then*

$$\liminf_{T\to\infty} \lambda_{\min} \left\{ \frac{1}{T} \sum_{t=1}^{T} \begin{pmatrix} \mathbf{x}_t \\ 1 \\ t/T \end{pmatrix}^{\otimes 2} \right\} \overset{a.s.}{>} 0, \tag{F.1}$$

and further for the non-explosive part, for all $\eta > 0$

$$\log \lambda_{\max} \left\{ \sum_{s=1}^{T} \begin{pmatrix} v_t \\ 1 \\ t \end{pmatrix} \begin{pmatrix} v_t \\ 1 \\ t \end{pmatrix}' \right\} \overset{a.s.}{=} \mathrm{o}\left(T^{\eta}\right) \tag{F.2}$$

$$\lim_{T\to\infty} \max_{t\le T} \begin{pmatrix} v_t \\ 1 \\ t \end{pmatrix}' \left\{ \sum_{s=1}^{T} \begin{pmatrix} v_t \\ 1 \\ t \end{pmatrix} \begin{pmatrix} v_t \\ 1 \\ t \end{pmatrix}' \right\}^{-1} \begin{pmatrix} v_t \\ 1 \\ t \end{pmatrix} \overset{a.s.}{=} 0. \tag{F.3}$$

Proof of Lemma F.1. (F.1): By a linear transformation the last component can be replaced by $(t - \sum_{s=1}^{T} s/T)/T = \tilde{t}$. The proof can be given along the lines of Chan and Wei (1985, Theorem 3). That is, let

$$\det\left(G_k - I_{pk+p}\right) = \sum_{j=0}^{pk+k} a_j \lambda^{pk+p-j}$$

be the characteristic polynomial of $\mathbf{x}_t$ with $a_0 = 1$. Define

$$y_t = \mathbf{x}_t + \sum_{j=1}^{pk+p} a_j \mathbf{x}_{t-j}, \qquad t > pk + p.$$

By the Cayley Hamilton Theorem $\sum_{j=0}^{pk+k} a_j G_k^{pk+p-j} = 0$ whereas by the model equation

$$\mathbf{x}_{t-j} = \sum_{l=0}^{pk+p-j-1} G_k^j \mathbf{e}_{t-j-l} + G_k^{pk+p} \mathbf{x}_{t-pk-p}, \qquad \text{for } j = 0, \ldots, pk + p - 1,$$

and therefore

$$y_t = \sum_{j=0}^{pk+p-1} \left(\sum_{l=0}^{j} a_l G_k^l \right) \mathbf{e}_{t-j}.$$

Using this formula Chan and Wei (1985, Theorem 3) show that $T^{-1}\sum_{t=pk+p+1}^{T} y_t y_t'$ is relatively compact with probability one and that $\liminf \lambda_{\min}(T^{-1}\sum_{t=pk+p+1}^{T} y_t y_t') \overset{a.s.}{>} 0$. Now, it will be shown that corresponding results hold for the extended vector $\mathbf{y}_t = (y_t', 1, \tilde{t})'$. Note, that $T^{-1}\sum_{t=pk+p+1}^{T} y_t\,(1, t/T) \overset{a.s.}{=} \mathrm{o}(1)$ since $T^{-1+d}\sum_{t=pk+p+1}^{T} t^d \mathbf{e}_t \overset{a.s.}{=} \mathrm{o}(1)$ by the Law of Large Numbers, see Hall and Heyde (1980, Theorem 2.18) and therefore $T^{-1}\sum_{t=pk+p+1}^{T} \mathbf{y}_t\mathbf{y}_t'$ has the same set of limit points as the diagonal matrix $\mathrm{diag}(T^{-1}\sum_{t=pk+p+1}^{T} y_t y_t', 1, 1/12)$. In particular, $\liminf \lambda_{\min}(T^{-1}\sum_{t=pk+p+1}^{T} \mathbf{y}_t\mathbf{y}_t') \overset{a.s.}{>} 0$. Therefore, following the remainder of the proof of Lai and Wei, it suffices to show that

$$\lambda_{\min}\left(\sum_{t=pk+p+1}^{T} \mathbf{y}_t\mathbf{y}_t'\right) \leq (pk+p+1)\left(\sum_{j=0}^{pk+p} a_j^2\right)\lambda_{\min}\left\{\sum_{t=1}^{T}\begin{pmatrix}\mathbf{x}_t\\ 1\\ \tilde{t}\end{pmatrix}^{\otimes 2}\right\}.$$

Let u be vector with length one. Then

$$u'\mathbf{y}_t\mathbf{y}_t'u = \left\{\sum_{j=0}^{pk+p} a_j u'\begin{pmatrix}\mathbf{x}_t\\ 1\\ \tilde{t}\end{pmatrix}\right\}^2 \leq \left(\sum_{j=0}^{pk+p} a_j^2\right)\sum_{j=0}^{pk+p}\left\{u'\begin{pmatrix}\mathbf{x}_t\\ 1\\ \tilde{t}\end{pmatrix}\right\}^2$$

and therefore

$$\begin{aligned}\sum_{t=pk+p+1}^{T} u'\mathbf{y}_t\mathbf{y}_t'u &\leq \left(\sum_{j=0}^{pk+p} a_j^2\right)\sum_{t=pk+p+1}^{T}\sum_{j=0}^{pk+p}\left\{u'\begin{pmatrix}\mathbf{x}_t\\ 1\\ \tilde{t}\end{pmatrix}\right\}^2\\ &\leq \left(\sum_{j=0}^{pk+p} a_j^2\right)(pk+p+1)\sum_{j=0}^{pk+p}\sum_{t=1}^{T} u'\begin{pmatrix}\mathbf{x}_t\\ 1\\ \tilde{t}\end{pmatrix}^{\otimes 2} u.\end{aligned}$$

The result then follows by ordering of positive semi-definite matrices.

(F.2): Note that

$$\begin{aligned}\sum_{t=1}^{T}\begin{pmatrix}v_t\\ 1\\ t\end{pmatrix}^{\otimes 2} &= \begin{pmatrix} I_{\dim V} & \sum_{t=1}^{T} v_t\,(1,t)\\ 0 & I_2\end{pmatrix}'\\ &\times\left\{\begin{array}{cc}\sum_{t=1}^{T}(v_t|1)^{\otimes 2} & 0\\ 0 & \sum_{t=1}^{T}\begin{pmatrix}1 & t\\ t & t^2\end{pmatrix}\end{array}\right\}\left\{\begin{array}{cc} I_{\dim V} & 0\\ \sum_{t=1}^{T}\begin{pmatrix}1\\ t\end{pmatrix} v_t' & I_2\end{array}\right\}.\end{aligned}$$

Thus by the definition of norms

$$\left\|\sum_{t=1}^{T}\begin{pmatrix}v_t\\ 1\end{pmatrix}^{\otimes 2}\right\| \leq \left\|\begin{pmatrix} I_{\dim V} & \sum_{t=1}^{T} v_t\,(1,t)\\ 0 & I_2\end{pmatrix}^{\otimes 2}\right\|\left\|\left\{\begin{array}{cc}\sum_{t=1}^{T}(v_t|1,t)^{\otimes 2} & 0\\ 0 & \sum_{t=1}^{T}\begin{pmatrix}1 & t\\ t & t^2\end{pmatrix}\end{array}\right\}\right\|.$$

For the first term on the right hand side

$$\begin{aligned}\left\|\begin{pmatrix} I_{\dim V} & \sum_{t=1}^{T} v_t\,(1,t) \\ 0 & I_2 \end{pmatrix}^{\otimes 2}\right\|^2 &\leq \operatorname{tr}\begin{pmatrix} I_{\dim V} & \sum_{t=1}^{T} v_t\,(1,t) \\ 0 & I_2 \end{pmatrix}^{\otimes 2} \\ &= 1+\dim V+\operatorname{tr}\left\{\sum_{t=1}^{T} v_t\,(1,t)\right\}^{\otimes 2} \overset{a.s.}{=} \mathrm{O}\left(T^{2\rho_1+4}\right)\end{aligned}$$

for some constant $\rho_1 > 0$ according to Lai and Wei (1985, Theorem 1) who prove that $||v_t|| \overset{a.s.}{=} \mathrm{O}(T^{\rho_1})$. For the second term use that

$$\begin{aligned}\left\|\begin{Bmatrix} \sum_{t=1}^{T} (v_t|1,t)^{\otimes 2} & 0 \\ 0 & \sum_{t=1}^{T}\begin{pmatrix} 1 & t \\ t & t^2 \end{pmatrix} \end{Bmatrix}\right\| &\leq \mathrm{O}\left(T^3\right)+\left\|\sum_{t=1}^{T}(v_t|1)^{\otimes 2}\right\| \\ &\leq \mathrm{O}\left(T^3\right)+\left\|\sum_{t=1}^{T} v_t^{\otimes 2}\right\| \overset{a.s.}{=} \mathrm{O}\left(T^{\rho_2}\right)\end{aligned}$$

for some constant $\rho_1 > 0$ according to Lai and Wei (1985, Corollary 1). The desired result then follows.

(F.3) The expression of interest can be decomposed as

$$\begin{aligned}&\begin{pmatrix} v_t \\ 1 \\ t \end{pmatrix}'\left\{\sum_{s=1}^{T}\begin{pmatrix} v_s \\ 1 \\ s \end{pmatrix}^{\otimes 2}\right\}^{-1}\begin{pmatrix} v_t \\ 1 \\ t \end{pmatrix} \\ =\;& \begin{pmatrix} 1 \\ t/T \end{pmatrix}'\left\{\sum_{s=1}^{T}\begin{pmatrix} 1 \\ s/T \end{pmatrix}^{\otimes 2}\right\}^{-1}\begin{pmatrix} 1 \\ t/T \end{pmatrix}+(v_t|1,t)'\left\{\sum_{s=1}^{T}(v_s|1,s)^{\otimes 2}\right\}^{-1}(v_t|1,t)\,.\end{aligned}$$

The first term is deterministic and of order o(1) since

$$\frac{1}{T}\sum_{s=1}^{T}\begin{pmatrix} 1 \\ s/T \end{pmatrix}^{\otimes 2}$$

has positive, convergent eigenvalues.

The second term can be shown to be of order o(1) by modifying Lai and Wei (1983, Theorem 4). That paper concerns a univariate autoregressive model without deterministic terms. The argument is as follows.

First, the inequality (C.1) in combination with Lemma C.1 imply

$$\liminf_{T\to\infty} \lambda_{\min}\frac{1}{T}\sum_{s=1}^{T}(v_s|1,s)^{\otimes 2} \overset{a.s.}{>} 0.$$

This together with Lai and Wei (1985, Theorem 1) can replace Lai and Wei (1983, Theorem 3) and Lai and Wei (1983, Lemmas 3-7) can then be modified.

Secondly, in the proof of Lai and Wei (1985, Theorem 4) use the modified Lemmas together with the following property. As a consequence of Lai and Wei (1982, equation 1.4b) then

$$\left\{\sum_{s=1}^{t}\begin{pmatrix} v_s \\ 1 \\ s \end{pmatrix}^{\otimes 2}\right\}^{-1} - \left\{\sum_{s=1}^{t+1}\begin{pmatrix} v_s \\ 1 \\ s \end{pmatrix}^{\otimes 2}\right\}^{-1} \geq 0,$$

and pre- and post-multiplication with $(I_{\dim V}, 0, 0)$ show

$$\left\{\sum_{s=1}^{t}(v_s|1,s)^{\otimes 2}\right\}^{-1} - \left\{\sum_{s=1}^{t+1}(v_s|1,s)^{\otimes 2}\right\}^{-1} \geq 0.$$

■

F.2 *Asymptotic results for* $\sum_{t=1}^{T} \mathbf{x}_{t-1}\mathbf{x}_{t-1}'$

This Section gives the details of the proof of Lemma D.1.

Lemma F.2 (Lemma D.1) *Suppose equation* (2.1), *the hypothesis,* $H(r)$, *condition* (i), *and Assumption 2.12 are satisfied. Then there exist normalisations* $N_a, N_{b_1}, \ldots, N_{b_l}, N_c, N_d$ *so*

$$\sum_{t=1}^{T} \{\operatorname{diag}(N_a, N_{b_1}, \ldots, N_{b_l}, N_c, N_d)\, M_k \mathbf{x}_k\}^{\otimes 2} \xrightarrow{\mathcal{D}} F = \operatorname{diag}(F_a, F_{b_1}, \ldots, F_{b_l}, F_c, F_d).$$

The limiting matrix F *is possibly random, but positive definite with probability one.*

The proof is given in the following sub-sections. First the Lemma C.2 is generalised to this situation in Section F.2.1. The convergence to F_a is discussed in Section F.2.2, to F_{b_j} in Sections F.2.3 and F.2.4, to F_c in Section F.2.5 and to F_d in Section F.2.6. The cross product terms are discussed in Section F.2.7.

F.2.1 *Generalising condition* (2.19)

Lemma F.3 *Suppose equation* (2.1), *condition* (i) *and Assumption 2.12 are satisfied. Then there exists a regular, real* $(pk+r)\times(pk+r)$*-matrix,* M*, so that*

$$MV_t = \left(MG_kM^{-1}\right)MV_{t-1} + M\eta_t$$

has the form

$$\begin{pmatrix} a_t \\ b_{1,t} \\ \vdots \\ b_{l,t} \\ c_t \\ d_t \end{pmatrix} = \begin{pmatrix} A & 0 & & \cdots & & 0 \\ 0 & B_1 & & & & \\ & & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & B_l & & \\ & & & & C & 0 \\ 0 & & \cdots & & 0 & D \end{pmatrix} \begin{pmatrix} a_{t-1} \\ b_{1,t-1} \\ \vdots \\ b_{l,t-1} \\ c_{t-1} \\ d_{t-1} \end{pmatrix} + \begin{pmatrix} \eta_{a,t} \\ \eta_{b_1,t} \\ \vdots \\ \eta_{b_l,t} \\ \eta_{c,t} \\ \eta_{d,t} \end{pmatrix},$$

where A *has roots at one,* B_j *has roots on the unit circle at the complex pairs* $\exp(i\theta_j)$ *and* $\exp(-i\theta_j)$ *where* $0<\theta_j\leq\pi$, C *has roots with modulus smaller than one and* D *has roots with modulus larger than one. Moreover,*

$$\liminf_T \left[E\left\{ \sum_{j=0}^{k} A^j \eta_{a,T+j}\eta'_{a,T+j}\left(A^j\right)' \middle| \mathcal{F}_T \right\} \right] \overset{a.s.}{>} 0, \tag{F.4}$$

$$\liminf_T \left[E\left\{ \sum_{j=0}^{k} B^j \eta_{b_j,T+j}\eta'_{b_j,T+j}\left(B^j\right)' \middle| \mathcal{F}_T \right\} \right] \overset{a.s.}{>} 0, \tag{F.5}$$

$$\liminf_T \left[E\left\{ \sum_{j=0}^{k} C^j \eta_{c,T+j}\eta'_{c,T+j}\left(C^j\right)' \middle| \mathcal{F}_T \right\} \right] \overset{a.s.}{>} 0, \tag{F.6}$$

$$\liminf_T \left[E\left\{ \sum_{j=0}^{k} D^j \eta_{d,T+j}\eta'_{d,T+j}\left(D^j\right)' \middle| \mathcal{F}_T \right\} \right] \overset{a.s.}{>} 0. \tag{F.7}$$

Note, that the sequence $\{M\eta_t\}$ *is a Martingale Difference.*

Proof of Lemma F.3. There exists a regular, real matrix which transforms G_k into a real, rational canonical form, see Herstein (1975, p.308). The rational canonical form is a special case of the suggested block diagonal matrix form MG_kM^{-1}. Further, in the Lemma C.2 it is proved that

$$\liminf_T \left[E\left\{ \sum_{j=0}^{k} G_k^j \eta_{T+j}\eta'_{T+j}\left(G_k^j\right)' \middle| \mathcal{F}_T \right\} \right] \overset{a.s.}{>} 0.$$

The results (F.4)-(F.7) then follow since M is regular and real. ∎

F.2.2 *Roots equal to one.*

In this subsection the process

$$a_t = A a_{t-1} + \eta_t$$

is considered, where A is a real square matrix and all its eigenvalues are equal to one. Further η_t satisfies the martingale difference Assumption 2.12 and the property (F.4) of Lemma F.3. First, the case where A is a Jordan block is considered and next the general result is given.

Lemma F.4 *Suppose A is a Jordan block of the form*

$$\begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & 0 \\ \vdots & & & \ddots & 1 \\ 0 & \cdots & \cdots & 0 & 1 \end{pmatrix}, \tag{F.8}$$

and that

$$\sum_{j=0}^{k} A^j E\left(\eta_{T+j}\eta'_{T+j}\middle|\mathcal{F}_T\right)\left(A^j\right)' \overset{a.s.}{>} 0. \tag{F.9}$$

Define the normalisation matrix

$$N_a = \begin{pmatrix} T^{\dim A} & & \\ & \ddots & \\ & & T \end{pmatrix}^{-1}.$$

Then $\sum_{t=1}^{T} N_a a_{t-1} a'_{t-1} N'_a \overset{\mathcal{D}}{\to} F_a$, where F_a is a stochastic matrix which is positive definite with probability one.

Proof of Lemma F.4. Partition a_t as $(a_{\dim A,t}, \ldots, a_{1,t})$ and η_t as $(\eta_{\dim A,t}, \ldots, \eta_{1,t})$, so that

$$\begin{aligned} a_{1,t} &= a_{1,0} + \sum_{s=1}^{t} \eta_{1,a,s}, \\ a_{j,t} &= a_{j,0} + \sum_{s=1}^{t} \eta_{j,a,s} + \sum_{s=1}^{t} a_{j-1,s-1} \qquad \text{for } j > 1. \end{aligned}$$

The j-power of the jordan block, A, is upper triangular with ones in the diagonal. It therefore follows from (F.4) of Lemma F.3 that also

$$\sum_{j=0}^{k} E\left(\eta_{1,T+j}\eta_{1,T+j}'\middle|\mathcal{F}_T\right) \overset{a.s.}{>} 0.$$

Since $\eta_{1,t}$ has constant variance it must be the case that $E(\eta_{1,t}^2|\mathcal{F}_{t-1})$ is positive and constant. In general, the other elements of η_t need not have positive variance.

An example is the bivariate case

$$E\left(\eta_{1,T+j}\eta_{1,T+j}'\middle|\mathcal{F}_T\right) = \begin{pmatrix} 0 & 0 \\ 0 & \sigma^2 \end{pmatrix}$$

in which case

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 0 & \sigma^2 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & \sigma^2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}\sigma^2$$

is positive definite as assumed. Let $\tilde{W}_1(u)$ be a univariate Brownian motion, with variance $E(\eta_{1,a,t}^2|\mathcal{F}_{t-1})$ and define recursively

$$\tilde{W}_j(v) = \int_0^v \tilde{W}_{j-1}(u)du.$$

Since $\tilde{\eta}_{1,a,t}$ is a martingale difference sequence with constant positive variance then

$$T^{-1/2}\sum_{s=1}^{[Tu]} \eta_{1,a,t} \overset{\mathcal{D}}{\to} \tilde{W}_1(u)$$

on $D[0,1]$, see Chan and Wei (1988, Theorem 2.4). Further for $j > 1$

$$T^{\eta-1/2}\sum_{s=1}^{[Tu]} \eta_{j,a,t} \overset{\mathcal{P}}{\to} 0$$

on $D[0,1]$ for all $\eta > 0$. Thus using the Continuous Mapping Theorem, see Billingsley (1968), repeatedly for the function from $D[0,1]$ into $D[0,1]$ defined by $x(v) \mapsto \int_0^v x(u)du$ it follows that

$$T^{1/2}N_a a_{[Tu]} \overset{\mathcal{D}}{\to} \left\{\tilde{W}_{\dim A}(u), \ldots, \tilde{W}_1(u)\right\}' \overset{\text{def}}{=} \tilde{W}(u)$$

on $D[0,1]^{\dim A}$ and

$$\sum_{t=1}^{T} N_a a_{t-1} a_{t-1} N_a' \overset{\mathcal{D}}{\to} F_a = \int_0^1 \tilde{W}(u)\,\tilde{W}'(u)\,du.$$

The stochastic matrix F_a is symmetric and positive definite with probability one, see Chan and Wei (1988, Lemma 3.11). ■

Lemma F.5 *Suppose A is a real square matrix with eigenvalues equal to one, and that the matrix*

$$\sum_{j=0}^{k} A^j E\left(\eta_{T+j}\eta'_{T+j}\middle|\mathcal{F}_T\right)\left(A^j\right)' \overset{a.s.}{>} 0. \tag{F.10}$$

Then there exists a normalisation matrix, N_a, so that $\sum_{t=1}^{T} N_a a_{t-1} a'_{t-1} N'_a \xrightarrow{\mathcal{D}} F_a$, where F_a is a stochastic matrix which is positive definite with probability one.

Proof. Since A has real roots there exists a real regular matrix, M_a, so that $M_a A M_a^{-1}$ is a real Jordan canonical form, see Herstein (1975, p.312). Thus $M_a A M_a^{-1}$ is a block matrix and for each block, $\tilde{A}_n$ say, and conformable blocks of η_t, $\tilde{\eta}_{n,t}$ say, it holds that

$$\sum_{j=0}^{k} \tilde{A}_n^j E\left(\tilde{\eta}_{n,t+j}\tilde{\eta}'_{n,t+j}\middle|\mathcal{F}_T\right)\left(\tilde{A}_n^j\right)' \overset{a.s.}{>} 0.$$

The result then follows from Lemma F.4. ■

F.2.3 *Roots equal to minus one*

Consider the process

$$b_t = Bb_{t-1} + \eta_t$$

where B is a real square matrix with all eigenvalues in minus one. The argument for the convergence $\sum_{t=1}^{T} N_b b_{t-1} b_{t-1}' N_b' \xrightarrow{\mathcal{D}} \mathcal{F}_b > 0$ is parallel to that of section F.2.2. See also Chan and Wei (1988, Section 3.2).

F.2.4 *Roots equal to* $\exp(i\theta_j)$ *for* $0 < \theta_j < 2\pi$.

In this subsection the process

$$b_t = Bb_{t-1} + \eta_t$$

is considered, where B is a real square matrix with all its roots in pairs, $\exp(i\theta)$ and $\exp(-i\theta)$ for $0 < \theta < \pi$. Further η_t satisfies the Martingale Difference Assumption 2.12 and the property (F.5).

Suppose the roots come in pairs $\exp(i\theta)$ and $\exp(-i\theta)$ for some $0 < \theta < \pi$. Before discussing the properties of the process in general, consider the case where b is a bivariate process, first in a lemma concerning the asymptotic properties of the process and next of the product moment matrix.

Lemma F.6 *Suppose the process* b_t,

$$B = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}. \tag{F.11}$$

and that $\operatorname{rank} E\left(\eta_{T+j}\eta_{T+j}' \middle| \mathcal{F}_T\right) \geq 1$. *Noting that*

$$B^t = \begin{Bmatrix} \cos(t\theta) & \sin(t\theta) \\ -\sin(t\theta) & \cos(t\theta) \end{Bmatrix} \tag{F.12}$$

it follows that for $u \in [0,1]$

$$\frac{1}{\sqrt{T}} \begin{bmatrix} \cos\{(\operatorname{int}(Tu)\theta)\} & -\sin\{(\operatorname{int}(Tu)\theta)\} \\ \sin\{(\operatorname{int}(Tu)\theta)\} & \cos\{(\operatorname{int}(Tu)\theta)\} \end{bmatrix} b_t \xrightarrow{\mathcal{D}} \sqrt{\frac{\sigma^2}{2}} W(u)$$

on $D[0,1]^2$ *where* W *is a standard bivariate Brownian motion and*

$$\sigma^2 = E\left\{\operatorname{tr}\left(\eta_{T+j}\eta_{T+j}'\right) \middle| \mathcal{F}_T\right\}.$$

Proof. First, the identity (F.12) follows since

$$B = \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} \begin{pmatrix} \exp(i\theta) & 0 \\ 0 & \exp(-i\theta) \end{pmatrix} \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}^{-1},$$

and hence

$$B^t = \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} \begin{pmatrix} \exp(i\theta) & 0 \\ 0 & \exp(-i\theta) \end{pmatrix}^t \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}^{-1}.$$

Using this expression and the definition $\eta_0 = b_0$, the process, b_t, can be written as

$$b_t = \sum_{s=0}^{t} B^{t-s}\eta_s = \begin{Bmatrix} \cos(t\theta) & \sin(t\theta) \\ -\sin(t\theta) & \cos(t\theta) \end{Bmatrix} \sum_{s=0}^{t} \begin{Bmatrix} \cos(s\theta) & -\sin(s\theta) \\ \sin(s\theta) & \cos(s\theta) \end{Bmatrix} \eta_s \qquad \text{(F.13)}$$

Secondly, to prove the asymptotic result partition η_t as $(\eta_{2,t}, \eta_{1,t})$. The cases when $E\left(\eta_{T+j}\eta'_{T+j}\middle| \mathcal{F}_T\right)$ has rank one and two are treated separately.

Suppose that $E\left(\eta_{T+j}\eta'_{T+j}\middle| \mathcal{F}_T\right) = \tilde{\Omega}$. Using Chan and Wei (1988, Theorem 2.2) and the Cramér-Wold theorem, see Billingsley (1968, p.49), it is seen that

$$\sqrt{\frac{2}{T}} \sum_{j=1}^{[Tu]} \left\{\eta_{1,t}\sin(j\theta), \eta_{2,t}\sin(j\theta), \eta_{1,t}\cos(j\theta), \eta_{2,t}\cos(j\theta)\right\}' \overset{\mathcal{D}}{\to} V_u$$

where V is a four dimensional Brownian motion with variance

$$\begin{pmatrix} \tilde{\Omega} & 0 \\ 0 & \tilde{\Omega} \end{pmatrix}.$$

The Continuous Mapping Theorem, see Billingsley (1968), then implies that

$$\sqrt{\frac{2}{T}} \sum_{j=1}^{[Tu]} \begin{Bmatrix} \cos(j\theta) & -\sin(j\theta) \\ \sin(j\theta) & \cos(j\theta) \end{Bmatrix} \begin{pmatrix} \tilde{\eta}_{2,j} \\ \tilde{\eta}_{1,j} \end{pmatrix} \overset{\mathcal{D}}{\to} \begin{pmatrix} V_{4,u} - V_{1,u} \\ V_{2,u} + V_{3,u} \end{pmatrix}.$$

The limiting Brownian motion can be represented as σW since

$$\begin{pmatrix} -1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \Omega & 0 \\ 0 & \Omega \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} = I_2 \mathrm{tr}\tilde{\Omega}.$$

■

Lemma F.7 *Suppose the process b_t is bivariate and that B is given by*

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}. \tag{F.14}$$

and that $\operatorname{rank}E\left(\eta_{T+j}\eta_{T+j}'\middle|\mathcal{F}_T\right) \geq 1$. *Let W be a standard bivariate Brownian motion and*

$$\sigma^2 = E\left\{\operatorname{tr}\left(\eta_{T+j}\eta_{T+j}'\right)\middle|\mathcal{F}_T\right\}.$$

The product moment matrix then converges to positive definite matrix proportional to the identity matrix,

$$\frac{1}{T^2}\sum_{t=1}^{T} b_t b_t' \xrightarrow{\mathcal{D}} \left\{\frac{\sigma^2}{4}\int_0^1 W_u' W_u du\right\} I_2. \tag{F.15}$$

Proof of Lemma F.7. Define the process

$$\begin{pmatrix} S_{2,t} \\ S_{1,t} \end{pmatrix} = \sum_{s=0}^{t}\left\{\begin{array}{cc} \cos(s\theta) & -\sin(s\theta) \\ \sin(s\theta) & \cos(s\theta) \end{array}\right\}\eta_t$$

using the convention $\eta_0 = b_0$. Thus by (F.12) it then follows that

$$b_t b_t' = \left\{\begin{array}{cc} \cos(t\theta) & \sin(t\theta) \\ -\sin(t\theta) & \cos(t\theta) \end{array}\right\}\begin{pmatrix} S_2^2 & S_1S_2 \\ S_1S_2 & S_1^2 \end{pmatrix}\left\{\begin{array}{cc} \cos(t\theta) & -\sin(t\theta) \\ \sin(t\theta) & \cos(t\theta) \end{array}\right\}.$$

By multiplying out it is seen that

$$\begin{aligned} b_t b_t' &= \begin{pmatrix} S_2^2 & S_1S_2 \\ S_1S_2 & S_1^2 \end{pmatrix}\cos^2(t\theta) + \begin{pmatrix} S_1^2 & -S_1S_2 \\ -S_1S_2 & S_2^2 \end{pmatrix}\sin^2(t\theta) \\ &\quad + \begin{pmatrix} 2S_1S_2 & S_1^2 - S_2^2 \\ S_1^2 - S_2^2 & -2S_1S_2 \end{pmatrix}\sin(t\theta)\cos(t\theta). \end{aligned}$$

Further, by the trigonometric identities $\cos(2\theta) = 2\cos^2\theta - 1 = 1 - 2\sin^2\theta$ and $2\sin\theta\cos\theta = \sin 2\theta$ it follows that

$$\begin{aligned} 2b_t b_t' &= \left(S_1^2 + S_2^2\right) I_2 + \begin{pmatrix} S_2^2 - S_1^2 & S_1S_2 \\ S_1S_2 & S_1^2 - S_2^2 \end{pmatrix}\cos(2t\theta) \\ &\quad + \begin{pmatrix} 2S_1S_2 & S_1^2 - S_2^2 \\ S_1^2 - S_2^2 & -2S_1S_2 \end{pmatrix}\sin(2t\theta). \end{aligned} \tag{F.16}$$

The normalised sum of the first term on the right hand side of this equation, (F.16), converges in distribution to (F.15), using Lemma F.6 and the Continuous Mapping Theorem for the integral of the outer product of a bivariate continuous function. Following Chan and Wei (1988, Lemma 3.3.6) the remaining terms are of order smaller than T^2 and can be ignored for asymptotic purposes. ■

Lemma F.8 *Suppose*

$$\left\{\begin{array}{ccccc} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} & I_2 & & & \\ & \ddots & \ddots & & \\ & & \ddots & I_2 & \\ & & & \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \end{array}\right\} \tag{F.17}$$

and that

$$\sum_{j=1}^{k} B^j E\left(\eta_{T+j}\eta_{T+j}' \middle| \mathcal{F}_T\right)\left(B^j\right)' \overset{a.s.}{>} 0.$$

Define the normalisation

$$N_b = \begin{pmatrix} T^{\dim B/2} I_2 & & \\ & \ddots & \\ & & T I_2 \end{pmatrix}.$$

Then $\sum_{t=1}^{T} N_b b_{t-1}^{\otimes 2} N_b \overset{\mathcal{D}}{\to} F_b$ *where* $F_b \overset{a.s.}{>} 0$.

Proof of Lemma F.8. The proof corresponds to that of Lemma F.4. Partition b_t as $(b_{\dim B,t}, \ldots, b_{1,t})'$ and η_t as $(\eta_{\dim B,t}, \ldots, \eta_{1,t})'$. Then

$$\begin{pmatrix} b_{2,t} \\ b_{1,t} \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}^t \begin{pmatrix} b_{2,0} \\ b_{1,0} \end{pmatrix} + \sum_{s=1}^{t} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}^{t-s} \begin{pmatrix} \eta_{2,s} \\ \eta_{1,s} \end{pmatrix}$$

and for $1 < j$

$$\begin{aligned} \begin{pmatrix} b_{2j,t} \\ b_{2j-1,t} \end{pmatrix} &= \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}^t \begin{pmatrix} b_{2j,0} \\ b_{2j-1,0} \end{pmatrix} \\ &\quad + \sum_{s=1}^{t} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}^{t-s} \left\{ \begin{pmatrix} \eta_{2j,s} \\ \eta_{2j-1,s} \end{pmatrix} + \begin{pmatrix} b_{2j-2,s-1} \\ b_{2j-3,s-1} \end{pmatrix} \right\} \end{aligned}$$

It follows that

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}^{-t} \begin{pmatrix} b_{2j,t} \\ b_{2j-1,t} \end{pmatrix} = \sum_{s=1}^{t} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}^{-s} \begin{pmatrix} b_{2j-2,s-1} \\ b_{2j-3,s-1} \end{pmatrix} + \mathrm{O}_{\mathcal{P}}\left(T^{1/2}\right).$$

Therefore let $\tilde{W}_1$ be a standard bivariate Brownian motion and for $j > 1$

$$\tilde{W}_j(v) = \int_0^v \tilde{W}_{j-1}(u)\, du.$$

Then

$$T^{1/2} N_b b_{[Tu]} \xrightarrow{\mathcal{D}} \sqrt{\frac{\sigma^2}{2}} \left\{ \tilde{W}'_{\dim B/2}(u), \ldots, \tilde{W}'_1(u) \right\}' \stackrel{def}{=} \sqrt{\frac{\sigma^2}{2}} \tilde{W}(u)$$

on $D[0,1]^{\dim B}$ and

$$\sum_{t=1}^{T} N_b b_{t-1}^{\otimes 2} N_b \xrightarrow{\mathcal{D}} F_b = \frac{\sigma^2}{2} \int_0^1 \tilde{W}(u) \tilde{W}(u) \, du.$$

■

Lemma F.9 *Suppose B is a real square matrix with eigenvalues equal to* $\exp(i\theta)$ *and* $\exp(-i\theta)$ *for some $0 < \theta < \pi$, and that the matrix*

$$\sum_{j=0}^{k} B^j E\left(\eta_{T+j} \eta'_{T+j} \middle| \mathcal{F}_T \right) \left(B^j \right)' \stackrel{a.s.}{>} 0.$$

Then there exists a normalisation matrix, N_b, so that $\sum_{t=1}^{T} N_b b_{t-1} b'_{t-1} N'_b \xrightarrow{\mathcal{D}} F_b$, where $F_b \stackrel{a.s.}{>} 0$.

Proof of Lemma F.9. There exist a regular matrix, M_b, so that $M_b B M_b^{-1}$ is block diagonal with blocks of the form (F.17). Since $M_b B M_b^{-1}$ is real then M_b can be chosen as a real matrix, see Herstein (1975, p.312). For each block, $\tilde{B}_n$ say, of $M_b B M_b^{-1}$ and comformable $\tilde{\eta}_{n,t}$ it holds

$$\sum_{j=0}^{k} \tilde{B}_n^j E\left(\tilde{\eta}_{n,T+j} \tilde{\eta}'_{n,T+j} \middle| \mathcal{F}_T \right) \left(\tilde{B}_n^j \right)' \stackrel{a.s.}{>} 0.$$

and the Lemma F.8 can be applied. ■

F.2.5 *Roots with modulus smaller than one*

The result follows from Lai and Wei (1985, Theorem 2, equation 3.9) using the martingale difference assumption and the Lemma C.2. They show that the normalisation can be chosen as $N_c = T^{-1/2}$. The limit, F_c, is deterministic.

F.2.6 *Roots with modulus greater than one*

The result follows from Lai and Wei (1985, Corollary 2) using the martingale difference assumption and the Lemma C.2. They show that the normalisation can be chosen as $N_d = D^{-T}$ and that $N_d \sum_{t=1}^{T} d_{t-1} d_{t-1}' N_d'$ converges with probability one to a random, prositive definite matrix, F_d.

F.2.7 *Cross product terms*

This section contains three Lemmae. The first covers cross products between two terms with distinct roots on the unit circle, the next discusses cross products between terms with roots on the unit circle and terms with stationary roots, and finally cross products between explosive and non-explosive terms are discussed.

Lemma F.10

$$N_{b_j} \sum_{t=1}^{T} b_{j,t-1} \left(a_t' N_a', b_{k,t-1}' N_{b_k}' \right) \xrightarrow{\mathcal{P}} 0 \qquad \text{for } j \neq k.$$

Proof of Lemma F.10. Chan and Wei (1988, Theorem 3.4.1) prove the result for a univariate process, $\dim X_t = 1$. Combining that result with the Jordan type decompositions in the Sections F.2.2-F.2.4 show that the multivariate results also holds. ∎

Lemma F.11

$$N_c \sum_{t=1}^{T} c_{t-1} \left(a_t' N_a', b_{j,t-1}' N_{b_j}' \right) \xrightarrow{\mathcal{P}} 0.$$

Proof of Lemma F.11. The argument correspond to that in the proof of Lemma F.10 with the only difference being that Chan and Wei (1988, Theorem 3.4.2) is used. ∎

Lemma F.12 *Define*

$$u_t = \left(a_t', b_{1,t}', \ldots, b_{l,t}', c_t' \right)' \qquad \text{and} \qquad N_u = diag(N_a, N_{b_1}, \ldots, N_{b_l}, N_c).$$

Then

$$N_d \sum_{t=1}^{T} d_{t-1} u_{t-1}' N_u \overset{a.s.}{\to} 0.$$

Proof of Lemma F.12. This follows from the proof of Lemma A.5. That is, by Lai and Wei (1985, Lemma 4 and Corollary 4)

$$\max_{1 \le t \le T} u_t' \left(\sum_{t=1}^{T} u_{t-1} u_{t-1}' \right)^{-1} u_t \overset{a.s.}{\to} 0, \qquad \text{and} \qquad \lim_{T \to \infty} \sum_{t=1}^{T} \|N_d d_t\| \overset{a.s.}{<} \infty.$$

Thus following the proof of Lai and Wei (1983, Theorem 1)

$$\begin{aligned} & \left\| \sum_{t=1}^{T} \left(\sum_{s=1}^{T} u_{t-1} u_{t-1}' \right)^{-1/2} u_{t-1} d_{t-1}' N_d' \right\| \\ \le & \max_{1 \le t \le T} \left\| \left(\sum_{s=1}^{T} u_{t-1} u_{t-1}' \right)^{-1/2} u_{t-1} \right\| \sum_{t=1}^{T} \left\| N_d d_{t-1}' \right\| \overset{a.s.}{\to} 0. \end{aligned}$$

■

F.3 *Asymptotic results for* $(\sum_{t=1}^{T} u_{t-1}u'_{t-1})^{-1/2} \sum_{t=1}^{T} u_{t-1}y'_t$

This section gives the details of the proof of Lemma D.2

Lemma F.13 (Lemma D.2) *Suppose equation* (2.1)*, the hypothesis,* $H(r)$*, condition* (i)*, and Assumption 2.12 are satisfied. Let* u_t *be given as either of the processes* $b_{j,t}, c_t, d_t$ *while* y_t *is either* $(t/T)^k$ *or* $T^{-k-1/2}\sum_{s=1}^{t-1}(t-s)^k\varepsilon_s$ *for* $k = 0, 1, \ldots$ *Then, for all* $\eta > 0$

$$T^{-\eta}\left(\sum_{t=1}^{T} u_{t-1}u'_{t-1}\right)^{-1/2}\sum_{t=1}^{T} u_{t-1}y'_t \overset{\mathcal{P}}{\to} 0,$$

whereas

$$T^{-1/2}\left(\sum_{t=1}^{T} a_{t-1}a'_{t-1}\right)^{-1/2}\sum_{t=1}^{T} a_{j,t-1}y'_t = \mathrm{O}_{\mathcal{P}}(1).$$

The proof for the cases $b_{j,t}, c_t, d_t$ is essentially given in Nielsen (2000). That paper deals with a univariate autoregressive process, however, there are no substantive changes when generalising to the multivariate case.

The case a_t is not discussed in Nielsen (2000) but is a rather straightforward consequence of the analysis in Section F.4.5.

The details are given in the following subsections.

F.3.1 *When* u *has unit roots*

Lemma F.14 *Suppose equation* (2.1)*, the hypothesis,* $H(r)$*, condition* (i)*, and Assumption 2.12 are satisfied. Let* y_t *be either* $(t/T)^k$ *or* $T^{-k-1/2}\sum_{s=1}^{t-1}(t-s)^k\varepsilon_s$ *for* $k = 0, 1, \ldots$ *Then*

$$T^{-1/2}\left(\sum_{t=1}^{T} a_{t-1}a'_{t-1}\right)^{-1/2}\sum_{t=1}^{T} a_{j,t-1}y'_t = \mathrm{O}_{\mathcal{P}}(1).$$

Proof of Lemma F.14. It follows from the arguments in Section F.4.5 that a normalisation N_a can be found so that $T^{1/2}N_a a_{j,[Tu]}$ converges in distribution to a continuous function of a Brownian motion. Likewise $y_{[Tu]}$ converges in distribution to either a polynomial or a continuous function of a Brownian motion. It can be argued that the two Brownian motions converge jointly and the result then follows using the Continuous Mapping Theorem. ■

F.3.2 *When u has roots on unit circle, but no roots at one*

In this case the proof is based on a modified version of Chan and Wei (1988, Theorem 2.1, subsequent remark).

Lemma F.15 *Let $\{X_t\}$ be a sequence of random vectors so*
(i) $E||X_t|| = \mathrm{O}(t^\alpha)$ for some $\alpha > 0$,
(ii) there exist random variables $A_j(t)$, $B_j(s,t)$ and constants φ_j, ψ_j and c with $||X_t - X_s|| \leq \sum_{j=1}^{q} A_j(t) B_j(s,t)$,
(iii) $EA^2(t) = \mathrm{O}(t^{\varphi_j})$ and $EB_j^2 = \mathrm{O}\{t^{\psi_j}(t-s)\}$ for $t \geq s$.
If $2\alpha = \varphi_j + \psi_j + 1$ for all j and $\exp(i\theta) \neq 1$ then for all $\eta > 0$

$$\sup_{1\leq t\leq T} \left\| \sum_{t=1}^{T} \exp(i\theta s) X_s \right\| = \mathrm{o}_{\mathcal{P}}\left(T^{\alpha+\eta+1/2}\right).$$

Proof of Lemma F.15. The result follows by careful reading of Chan and Wei (1988, Theorem 2.1, subsequent remark). ■

Lemma F.16 *Suppose equation* (2.1)*, the hypothesis, $H(r)$, condition* (*i*)*, and Assumption 2.12 are satisfied. Let y_t be either $(t/T)^k$ or $T^{-k-1/2}\sum_{s=1}^{t-1}(t-s)^k \varepsilon_s$ for $k = 0, 1, \ldots$ Then, for all $\eta > 0$*

$$T^{-\eta} \left(\sum_{t=1}^{T} b_{j,t-1} b_{j,t-1}' \right)^{-1/2} \sum_{t=1}^{T} b_{j,t-1} y_t' \xrightarrow{\mathcal{P}} 0.$$

Proof of Lemma F.16. Find the Jordan forms as outlined in Sections F.2.3, F.2.4 and the corresponding normalisations N_b. It then suffices to prove

$$T^{-\eta} N_b \sum_{t=1}^{T} b_{j,t-1} y_t' \xrightarrow{\mathcal{P}} 0.$$

A typical element is the following. Let $k = 0$ and consider

$$\sum_{t=1}^{T} \left(\frac{1}{T} \sum_{s=1}^{t-1} (-1)^{t-s-1} \eta_s \right) \left(\frac{1}{T^{1/2}} \sum_{s=1}^{t-1} \varepsilon_s \right)' = \frac{1}{T^{3/2}} \sum_{t=1}^{T} (-1)^{t-1} X_t$$

where

$$X_t = \sum_{m=1}^{t-1} (-1)^m \eta_m \sum_{n=1}^{t-1} \varepsilon_n'.$$

Note that for $t \geq s$ then

$$\begin{aligned} X_t - X_s &= \sum_{m=1}^{t-1} (-1)^m \eta_m \sum_{n=s+1}^{t-1} \varepsilon_n' + \sum_{m=s+1}^{t-1} (-1)^m \eta_m \sum_{n=1}^{s} \varepsilon_n' \\ &\overset{def}{=} A_1(t) B_1(s,t) + B_2(s,t) A_2(t). \end{aligned}$$

Here $E||X_t|| = \mathrm{O}(t)$, $EA_j^2 = \mathrm{O}(t)$, $EB_j^2 = \mathrm{O}(t-s)$ showing $\alpha = 1$, $\varphi_j = 1$, $\psi_j = 0$ and $2\alpha = \varphi_j + \psi_j + 1 = 2$. Therefore Lemma F.15 show as desired that

$$\sup_{1 \leq t \leq T} \left\| \sum_{t=1}^{T} (-1)^{t-1} X_t \right\| = \mathrm{o}_{\mathcal{P}}\left(T^{3/2+\eta}\right).$$

Another typical element is

$$\sum_{t=1}^{T} \left(\frac{1}{T} \sum_{s=1}^{t-1} (-1)^{t-s-1} \eta_s \right) \frac{t^k}{T^k} = \frac{1}{T^{k+1}} \sum_{t=1}^{T} (-1)^{t-1} X_t$$

where

$$X_t = \sum_{m=1}^{t-1} (-1)^m \eta_m t^{k-1} \sum_{n=1}^{t-1} 1.$$

Here $E||X_t|| = \mathrm{O}(t^{k+1/2})$, $EA_1^2 = \mathrm{O}(t)$, $EB_1^2 = \mathrm{O}\{t^{2k-2}(t-s)^2\}$, $EA_2^2 = \mathrm{O}(t^{2k})$, $EB_2^2 = \mathrm{O}(t-s)$ showing $\alpha = k + 1/2$, $\varphi_1 = 1$, $\psi_1 = 2k - 1$, $\varphi_2 = 2k$, $\psi_2 = 0$, and $2\alpha = \varphi_j + \psi_j + 1 = 2k + 1$. Therefore Lemma F.15 show as desired that

$$\sup_{1 \leq t \leq T} \left\| \sum_{t=1}^{T} (-1)^{t-1} X_t \right\| = \mathrm{o}_{\mathcal{P}}\left(T^{k+1+\eta}\right).$$

■

F.3.3 When u has stationary roots

In this case the proof is based on a modified version of Chan and Wei (1988, Lemma 3.4.3).

Lemma F.17 *Assume $\{\mathbf{g}_t\}$ and $\{\mathbf{h}_t\}$ are two sequences of random vectors in $\mathbf{R}^s$ such that $\mathbf{g}_t$ and $\mathbf{h}_t$ are $\mathcal{F}_t$-measurable. Suppose there exists an $(s \times s)$ constant matrix M such that $\mathbf{g}_t = M\mathbf{g}_{t-1} + \mathbf{h}_t$, where $\mathbf{g}_0 = \mathbf{h}_0 = 0$. Further, suppose $E\sum_{t=1}^T ||\mathbf{g}_t||^2 = \mathrm{O}(T^\alpha)$ and $E\sum_{t=1}^T ||\mathbf{h}_t||^2 = \mathrm{O}(T^{\alpha-1})$ for some $\alpha \geq 2$. Next, let ε_t satisfy Assumption 2.12 and let c_t be given by $c_t = Cc_{t-1} + \eta_{c,t}$ as specified in Section F.2. Then for any fixed integer j and for all $\eta > 0$,*

$$E\left\|\sum_{t=1}^T \mathbf{g}_t \varepsilon'_{t+j}\right\| = \mathrm{O}\left(T^{\alpha/2}\right), \qquad E\left\|\sum_{t=1}^T \mathbf{g}_t c'_t\right\| = \mathrm{o}\left(T^{\alpha/2+\eta}\right)$$

Proof of Lemma F.17. The proof follows by careful reading of the proof of Chan and Wei (1988, Lemma 3.4.3). ■

Lemma F.18 *Suppose equation* (2.1)*, the hypothesis, $H(r)$, condition (i), and Assumption 2.12 are satisfied. Let y_t be either $(t/T)^k$ or $T^{-k-1/2}\sum_{s=1}^{t-1}(t-s)^k\varepsilon_s$ for $k = 0, 1, \ldots$ Then, for all $\eta > 0$*

$$T^{-\eta}\left(\sum_{t=1}^T c_{t-1}c'_{t-1}\right)^{-1/2}\sum_{t=1}^T c_{t-1}y'_t \overset{\mathcal{P}}{\to} 0.$$

Proof of Lemma F.18. As discussed in Section F.2.5 then $T^{-1}\sum_{t=1}^T c_{t-1}c'_{t-1} \overset{\mathcal{P}}{\to} F_c$ where $F_c \overset{a.s.}{>} 0$. Thus it suffices to argue that $T^{-1/2-\eta}\sum_{t=1}^T c_{t-1}y'_t \overset{\mathcal{P}}{\to} 0$.

The case where $y_t = T^{-k-1/2}\sum_{s=1}^{t-1}(t-s)^k\varepsilon_s$. Use Lemma F.17 with $\mathbf{g}_t = \sum_{s=1}^{t-1}(t-s)^k\varepsilon_s$ and $\alpha = 2k+2$. Then it follows that $\sum_{t=1}^T \mathbf{g}_t\mathbf{z}'_t \overset{\mathcal{L}_1}{=} \mathrm{o}(T^{k+1+\eta})$ and hence $\sum_{t=1}^T \mathbf{g}_t\mathbf{z}'_t \overset{\mathcal{P}}{=} \mathrm{o}(T^{k+1+\eta})$ as desired.

The case where $y_t = (t/T)^k$. The process c_t is a linear process with exponentially decreasing coefficients. For $k = 0$ the result then follows from the Central Limit Theorem for linear processes with martingale difference innovations, see Phillips and Solo (1992, Theorem 3.16). For general k the result follows by partial summation and the Central Limit Theorem together with the invariance principle also given in Phillips and Solo (1992, Theorem 3.16). ■

F.3.4 *When u has explosive roots*

Lemma F.19 *Suppose equation* (2.1)*, the hypothesis,* $H(r)$*, condition* (i)*, and Assumption 2.12 are satisfied. Let* y_t *be either* $(t/T)^k$ *or* $T^{-k-1/2}\sum_{s=1}^{t-1}(t-s)^k\varepsilon_s$ *for* $k=0,1,\ldots$ *Then, for all* $\eta>0$

$$T^{-\eta}\left(\sum_{t=1}^{T}d_{t-1}d_{t-1}'\right)^{-1/2}\sum_{t=1}^{T}d_{t-1}y_t' \overset{\mathcal{P}}{\to} 0.$$

Proof of Lemma F.19. As in the proof of Lemma A.5 then

$$\left\|\left(\sum_{t=1}^{T}d_{t-1}d_{t-1}'\right)^{-1/2}\sum_{t=1}^{T}d_{t-1}y_t'\right\| \overset{a.s.}{=} \left(\max_{t\leq T}\|y_t\|\right)\mathrm{O}\left(1\right)$$

and it therefore suffices to prove that $\max_{t\leq T}\|y_t\| = \mathrm{O}_{\mathcal{P}}\left(1\right).$

The case where $y_t = T^{-k-1/2}\sum_{s=1}^{t-1}(t-s)^k\varepsilon_s$. In this situation $y_{[Tu]}$ converges in distribution on $D[0,1]^p$ but has continouous paths. The maximum function is continuous on $C[0,1]$ on therefore $\max_{t\leq T}\|y_t\| = \mathrm{O}_{\mathcal{P}}\left(1\right).$

The case where $y_t=(t/T)^k$. Then $\max_{t\leq T}\|y_t\|=1.$ ■

F.4 *Generalisation to the model with constant term*

Throughout this Section it is assumed that $\beta_c = 0$ in the probabilistic analysis following Theorem 2.9, whereas $\beta_c' \in \mathbf{R}^r$ in the statistical analysis.

F.4.1 *Consistency of eigenvalues*

This argument generalises the proof of Theorem 2.4. The argument is simply extended drawing on (A.5) of Lemma A.7.

Lemma F.20 *Suppose the model is given by the equation* (2.6), *the Assumption 2.12, the hypothesis* $H_c(r)$ *and the condition* (*i*) *are satisfied. Then*

$$\hat{\lambda}_p^c, \ldots, \hat{\lambda}_{r+1}^c \overset{a.s.}{\to} 0, \qquad \liminf_{T\to\infty} \hat{\lambda}_r^c > 0.$$

Proof of Lemma F.20. For notational simplicity it is assumed that $k = 0$. In this case the dual eigenvalue problem is

$$\begin{aligned} 0 &= \det\left[\lambda \sum_{t=1}^T \Delta X_t^{\otimes 2} - \sum_{t=1}^T \Delta X_t \begin{pmatrix} X_{t-1} \\ 1 \end{pmatrix}' \right. \\ &\qquad \left. \times \left\{\sum_{t=1}^T \begin{pmatrix} X_{t-1} \\ 1 \end{pmatrix}^{\otimes 2}\right\}^{-1} \sum_{t=1}^T \begin{pmatrix} X_{t-1} \\ 1 \end{pmatrix} \Delta X_t' \right] \\ &= \det\left[\lambda S_{00} - S_{01} S_{11}^{-1} S_{10} - \sum_{t=1}^T \Delta X_t \left(1 | X_{t-1}\right) \left\{\sum_{t=1}^T \left(1|X_{t-1}\right)^2\right\} \sum_{t=1}^T \left(1|X_{t-1}\right) \Delta X_t\right] \end{aligned}$$

Using (A.5) of Lemma A.7 and the boundedness of $S_{\beta\beta}^{-1}$ given by Lemma A.4 it follows that

$$A_T' \sum_{t=1}^T \Delta X_t \left(1|X_{t-1}\right) \left\{\sum_{t=1}^T \left(1|X_{t-1}\right)^2\right\} \sum_{t=1}^T \left(1|X_{t-1}\right) \Delta X_t A_T \overset{a.s.}{\to} 0.$$

The argument then follows as in the proof of Theorem 2.4. ■

F.4.2 *Asymptotic distribution of rank rest*

This argument generalises the proof of Theorem 2.5.

Lemma F.21 *Suppose equation* (2.6), *the Assumption 2.12, hypothesis* $H_c(r)$ *and the conditions* (i), (ii) *are satisfied. Then*

$$
\begin{aligned}
& LR\{H_c(r)|\, H_c(p)\} \\
& \xrightarrow{\mathcal{D}} tr\left[\int_0^1 dW_u \begin{pmatrix} W_u \\ 1 \end{pmatrix}' \left\{\int_0^1 \begin{pmatrix} W_u \\ 1 \end{pmatrix}\begin{pmatrix} W_u \\ 1 \end{pmatrix}' du\right\}^{-1} \int_0^1 \begin{pmatrix} W_u \\ 1 \end{pmatrix} dW_u'\right].
\end{aligned}
$$

Proof of Lemma F.21. For notational simplicity assume that $k = 0$. Replace the matric C_T by $C_{T,c} = \mathrm{diag}(C_T, 1)$. Then the eigenvalue problem is

$$
\begin{aligned}
0 &= \det C_{T,c}' \left\{ \frac{\lambda}{T} \sum_{t=1}^{T} \begin{pmatrix} X_{t-1}|1 \\ 1 \end{pmatrix}^{\otimes 2} - \sum_{t=1}^{T} \begin{pmatrix} X_{t-1}|1 \\ 1 \end{pmatrix} \Delta X_t' \right. \\
& \qquad \left. \times \left(\sum_{t=1}^{T} \Delta X_t^{\otimes 2}\right)^{-1} \sum_{t=1}^{T} \Delta X_t \begin{pmatrix} X_{t-1}|1 \\ 1 \end{pmatrix}' \right\} C_{T,c}.
\end{aligned}
$$

Because of the positive definiteness of $\sum_{t=1}^{T} (\beta' X_{t-1}|1)$ then

$$
\left\| \left\{ \sum_{t=1}^{T} (\beta' X_{t-1})^{\otimes 2} \right\}^{-1/2} \sum_{t=1}^{T} (\beta' X_{t-1}|1)^{\otimes 2} \left\{ \sum_{t=1}^{T} (\beta' X_{t-1})^{\otimes 2} \right\}^{-1/2} \right\| \leq 1
$$

and it is equivalent to considet

$$
\begin{aligned}
0 &= \det \left\{ \frac{\lambda}{T} \begin{pmatrix} I_{p-r} & & \\ & 0 & \\ & & 1 \end{pmatrix} C_{T,c}' \sum_{t=1}^{T} \begin{pmatrix} X_{t-1}|1 \\ 1 \end{pmatrix}^{\otimes 2} C_{T,c} \begin{pmatrix} I_{p-r} & & \\ & 0 & \\ & & 1 \end{pmatrix} \right. \\
& \qquad \left. - C_{T,c}' \sum_{t=1}^{T} \begin{pmatrix} X_{t-1}|1 \\ 1 \end{pmatrix} \Delta X_t' \left(\sum_{t=1}^{T} \Delta X_t^{\otimes 2}\right)^{-1} \sum_{t=1}^{T} \Delta X_t \begin{pmatrix} X_{t-1}|1 \\ 1 \end{pmatrix}' \right\} C_{T,c}.
\end{aligned}
$$

Using Corollary A.8 in the same way as in the proof of Theorem 2.5 this leads to

$$
0 = \det \alpha_\perp' \left\{ \lambda \int_0^1 \begin{pmatrix} W_u \\ 1 \end{pmatrix}^{\otimes 2} du - \int_0^1 \begin{pmatrix} W_u \\ 1 \end{pmatrix} dW_u' \alpha_\perp (\alpha_\perp' \Omega \alpha_\perp)^{-1} \alpha_\perp' \int_0^1 dW_u \begin{pmatrix} W_u \\ 1 \end{pmatrix} \right\} \alpha_\perp.
$$

and the result follows. ■

F.4.3 *Consistency of* $\tilde{\psi}_\perp$

The Lemma A.9 has to be generalised. Let $\tilde{\psi}_\perp$ be estimated under the model with a constant (2.6). Suppose $\alpha = A\psi$ and $\beta = H\varphi$. For some random $\{r \times (m-r)\}$-matrix U_T let

$$\tilde{\psi}_\perp = \hat{\psi}_\perp \left(\overline{\psi}_\perp' \hat{\psi}_\perp\right)^{-1} = \psi_\perp + \overline{\psi}\psi'\tilde{\psi}_\perp \overset{\text{def}}{=} \psi_\perp + \overline{\psi}U_T.$$

Lemma F.22 *Let* $\tilde{\psi}_\perp$ *be estimated under the model with a constant* (2.6)*. Suppose equation* (2.1)*, the hypotheses* $H_\alpha(r)$*,* $H_\beta(r)$*, condition* (*i*) *and Assumption* 2.12 *are satisfied. Let* $\overline{A}_\omega = \overline{A} - A_\perp \Omega_{A_\perp A_\perp}^{-1} \Omega_{A_\perp A}$*. Then, for all* $\xi < \gamma/(2+\gamma)$

$$S_{\beta\beta}^{1/2} U_T \overset{a.s.}{=} \mathrm{o}\left(T^{-\xi/2}\right), \tag{F.18}$$

$$\tilde{\psi}_\perp \overset{a.s.}{=} \psi_\perp + \mathrm{o}\left(T^{-\xi/2}\right), \tag{F.19}$$

$$-S_{\beta\beta}^{1/2} U_T = S_{\beta\beta}^{-1/2} S_{\beta\varepsilon} \overline{A}_\omega \psi_\perp \left\{1 + \mathrm{o}_\mathcal{P}\left(T^{-\xi/2}\right)\right\} + \mathrm{o}_\mathcal{P}\left(T^{-\xi}\right). \tag{F.20}$$

Proof of Lemma F.22. It is convenient to define $M_\psi = (\psi_\perp, \overline{\psi} S_{\beta\beta}^{-1/2})$ and recall the statements (C.8)-(C.10)

$$M_\psi' S_{AA\cdot A_\perp} M_\psi \overset{a.s.}{=} M_\psi' \Omega_{AA\cdot A_\perp} M_\psi + \begin{pmatrix} 0 & 0 \\ 0 & I_r \end{pmatrix} + \mathrm{o}\left(T^{-\xi/2}\right) \tag{F.21}$$

$$S_{HH\cdot A_\perp} \overset{a.s.}{=} S_{HH}\left\{1 + \mathrm{o}\left(T^{-\xi}\right)\right\}, \tag{F.22}$$

$$M_\psi' S_{AH\cdot A_\perp} S_{HH\cdot A_\perp}^{-1} S_{HA\cdot A_\perp} M_\psi \overset{a.s.}{=} \begin{pmatrix} 0 & 0 \\ 0 & I_r \end{pmatrix} + \left\{\begin{matrix} \mathrm{o}\left(T^{-\xi}\right) & \mathrm{o}\left(T^{-\xi/2}\right) \\ \mathrm{o}\left(T^{-\xi/2}\right) & \mathrm{o}\left(T^{-\xi/2}\right) \end{matrix}\right\} \tag{F.23}$$

Further, by Lemmas A.5, A.6 and (F.21)

$$S_{\varepsilon H\cdot A_\perp} S_{HH}^{-1/2} \overset{a.s.}{=} \mathrm{o}\left(T^{-\xi/2}\right), \tag{F.24}$$

(A.4), (A.5) of Lemma A.7

$$S_{cc\cdot H,A_\perp} \overset{a.s.}{=} S_{cc\cdot H} + \mathrm{o}\left(T^{-\xi}\right), \tag{F.25}$$

and by Lemma A.4 and (A.1), (A.4), (A.5) of Lemma A.7

$$M_\psi' S_{Ac\cdot H,A_\perp} S_{cc\cdot H}^{-1/2} = M_\psi' \left(\overline{A}' - S_{AA_\perp\cdot H} S_{A_\perp A_\perp\cdot H}^{-1} A_\perp'\right) S_{0c\cdot H_\perp} S_{cc\cdot H}^{-1/2} \overset{a.s.}{=} \mathrm{o}\left(T^{-\xi/2}\right). \tag{F.26}$$

(F.18): The concentrated model equation (2.6) is

$$R_{0,t} = A\psi\left(\varphi' H' R_{1,t} + \beta_c R_{c,t}\right) + \hat{\varepsilon}_t. \tag{F.27}$$

Pre-multiplication by $(\overline{A}, A'_{\perp})$ shows that $A'_{\perp}X_t$ is weakly exogeneous for ψ, φ, β_c. Thus likelihood inference can be based on the partial system of $\overline{A}'X_t$ given $A'_{\perp}X_t$, that is

$$\overline{A}'R_{0,t} = \psi\left(\varphi' H' R_{1,t} + \beta_c R_{c,t}\right) + \omega A'_{\perp} R_{0,t} + \overline{A}'_{\omega}\hat{\varepsilon}_t. \tag{F.28}$$

for $\overline{A}'_{\omega} = (\overline{A}' - \omega A'_{\perp})$ and $\omega = \Omega_{AA_{\perp}}\Omega^{-1}_{A_{\perp}A_{\perp}}$. The squared sample canonical correlations for $\overline{A}'R_{0,t}$ and $(R'_{1,t}H, R_{c,t})$ solve

$$0 = \det M'_{\psi}\left(\lambda S_{AA\cdot A_{\perp}} - S_{AH\cdot A_{\perp}} S^{-1}_{HH\cdot A_{\perp}} S_{HA\cdot A_{\perp}} - S_{Ac\cdot H,A_{\perp}} S^{-1}_{cc\cdot H,A_{\perp}} S_{cA\cdot H,A_{\perp}}\right) M_{\psi}$$

The equations (F.25), (F.26) show that equivalently

$$\begin{aligned} 0 &\overset{a.s.}{=} \det M'_{\psi}\left\{\lambda S_{AA\cdot A_{\perp}} - S_{AH\cdot A_{\perp}} S^{-1}_{HH\cdot A_{\perp}} S_{HA\cdot A_{\perp}} + \mathrm{o}\left(T^{-\xi}\right)\right\} M_{\psi} \\ &\overset{a.s.}{=} \det\left\{P_{\psi}(\lambda) + \mathrm{o}\left(T^{-\xi/2}\right)\right\} \end{aligned}$$

where (C.8), (C.10) have been used and

$$P_{\psi}(\lambda) = \lambda M'_{\psi}\Omega_{AA\cdot A_{\perp}} M_{\psi} + (\lambda - 1)\begin{pmatrix} 0 & 0 \\ 0 & I_r \end{pmatrix}.$$

For $\lambda = 0$ then $(I_{m-r}, 0)P_A(\lambda) = 0$ which shows that

$$\operatorname{span}\left\{\begin{pmatrix} \overline{\psi}'_{\perp} \\ S^{1/2}_{\beta\beta}\psi' \end{pmatrix}\hat{\psi}_{\perp}\right\} \overset{a.s.}{=} \operatorname{span}\begin{pmatrix} I_{m-r} \\ 0 \end{pmatrix} + \mathrm{o}\left(T^{-\xi/2}\right). \tag{F.29}$$

Since $\overline{\psi}'_{\perp}\hat{\psi}_{\perp}$ has full rank almost surely then it also holds that

$$\operatorname{span}\left\{\begin{pmatrix} \overline{\psi}'_{\perp} \\ S^{1/2}_{\beta\beta}\psi' \end{pmatrix}\hat{\psi}_{\perp}\right\} \overset{a.s.}{=} \operatorname{span}\begin{pmatrix} I_{m-r} \\ 0 \end{pmatrix} + \mathrm{o}\left(T^{-\xi/2}\right).$$

(A.8): By (A.6) it holds $\tilde{\psi}_{\perp} = \psi_{\perp} + \overline{\psi} S^{-1/2}_{\beta\beta} S^{1/2}_{\beta\beta}\overline{\psi}'_{\perp}\hat{\psi}_{\perp}$ and the result follows by the boundedness of $S^{-1/2}_{\beta\beta}$, see Lemma A.4.

(A.9): The partial likelihood given by (F.28) is differentiated with respect to ψ in the direction a following Johansen (1996, equation 13.9). This gives the likelihood equation

$$\begin{aligned} 0 &= \frac{\partial}{\partial\psi}\log L(\psi, \varphi, \Omega) \\ &= T\operatorname{tr}\left[\Omega^{-1}_{AA\cdot A_{\perp}}\left\{(S_{AH\cdot A_{\perp}}, S_{Ac\cdot A_{\perp}}) - \psi(\varphi', \beta'_c)\begin{pmatrix} S_{HH\cdot A_{\perp}} & S_{Hc\cdot A_{\perp}} \\ S_{cH\cdot A_{\perp}} & S_{cc\cdot A_{\perp}} \end{pmatrix}\right\}\begin{pmatrix} \varphi \\ \beta_c \end{pmatrix} a'\right]. \end{aligned}$$

In particular it holds that

$$S_{AH\cdot A_\perp}\begin{pmatrix}\hat{\varphi}\\ \hat{\beta}_c\end{pmatrix}=\hat{\psi}\left(\hat{\varphi}',\hat{\beta}_c'\right)\begin{pmatrix}S_{HH\cdot A_\perp} & S_{Hc\cdot A_\perp}\\ S_{cH\cdot A_\perp} & S_{cc\cdot A_\perp}\end{pmatrix}\begin{pmatrix}\hat{\varphi}\\ \hat{\beta}_c\end{pmatrix}.$$

Since $\tilde{\psi}_\perp'\hat{\psi}=0$ and ψ has full column rank by condition (i) then

$$0=\tilde{\psi}_\perp'\left(S_{AH\cdot A_\perp},S_{Ac\cdot A_\perp}\right)\begin{pmatrix}\hat{\varphi}\\ \hat{\beta}_c\end{pmatrix}\hat{\psi}'\overline{\psi}.$$

Following Johansen (1996, p. 130) the maximum likelihood estimator for (φ',β_c') is

$$\begin{pmatrix}\hat{\varphi}\\ \hat{\beta}_c\end{pmatrix}=\begin{pmatrix}S_{HH\cdot\hat{\alpha}_\perp} & S_{Hc\cdot\hat{\alpha}_\perp}\\ S_{cH\cdot\hat{\alpha}_\perp} & S_{cc\cdot\hat{\alpha}_\perp}\end{pmatrix}^{-1}\begin{pmatrix}S_{HA\cdot\hat{\alpha}_\perp}\\ S_{cA\cdot\hat{\alpha}_\perp}\end{pmatrix}\hat{\psi}\left(\hat{\psi}'\hat{\psi}\right)^{-1}$$

and therefore

$$\begin{aligned}0 &= \tilde{\psi}_\perp'\left(S_{AH\cdot A_\perp},S_{Ac\cdot A_\perp}\right)\begin{pmatrix}S_{HH\cdot\hat{\alpha}_\perp} & S_{Hc\cdot\hat{\alpha}_\perp}\\ S_{cH\cdot\hat{\alpha}_\perp} & S_{cc\cdot\hat{\alpha}_\perp}\end{pmatrix}^{-1}\begin{pmatrix}S_{HA\cdot\hat{\alpha}_\perp}\\ S_{cA\cdot\hat{\alpha}_\perp}\end{pmatrix}\hat{\psi}\left(\hat{\psi}'\hat{\psi}\right)^{-1}\hat{\psi}'\overline{\psi}\\ &\overset{a.s.}{=} \tilde{\psi}_\perp'\left(S_{AH\cdot A_\perp},S_{Ac\cdot A_\perp}\right)\begin{pmatrix}S_{HH\cdot\hat{\alpha}_\perp} & S_{Hc\cdot\hat{\alpha}_\perp}\\ S_{cH\cdot\hat{\alpha}_\perp} & S_{cc\cdot\hat{\alpha}_\perp}\end{pmatrix}^{-1}\begin{pmatrix}S_{HA\cdot\hat{\alpha}_\perp}\\ S_{cA\cdot\hat{\alpha}_\perp}\end{pmatrix}\overline{\psi}\left\{1+\mathrm{o}\left(T^{-\xi/2}\right)\right\},\end{aligned}$$

where the consistency of $\tilde{\psi}_\perp$ given in (F.19) has been used. Using the concentrated model equation (F.27) this implies

$$\begin{aligned}0 &\overset{a.s.}{=} \tilde{\psi}_\perp'\left\{\psi\varphi'\left(S_{HH\cdot A_\perp},S_{Hc\cdot A_\perp}\right)+\overline{A}'\left(S_{\varepsilon H\cdot A_\perp},S_{\varepsilon c\cdot A_\perp}\right)\right\}\begin{pmatrix}S_{HH\cdot\hat{\alpha}_\perp} & S_{Hc\cdot\hat{\alpha}_\perp}\\ S_{cH\cdot\hat{\alpha}_\perp} & S_{cc\cdot\hat{\alpha}_\perp}\end{pmatrix}^{-1}\\ &\quad\times\left\{\begin{pmatrix}S_{HH\cdot\hat{\alpha}_\perp}\\ S_{cH\cdot\hat{\alpha}_\perp}\end{pmatrix}\varphi\psi'+\begin{pmatrix}S_{H\varepsilon\cdot\hat{\alpha}_\perp}\\ S_{c\varepsilon\cdot\hat{\alpha}_\perp}\end{pmatrix}\right\}\overline{\psi}\left\{1+\mathrm{o}\left(T^{-\xi/2}\right)\right\}.\end{aligned}$$

which is equivalent to

$$\begin{aligned}0 &\overset{a.s.}{=} \tilde{\psi}_\perp'\left\{\psi\varphi'\left(S_{HH\cdot A_\perp},S_{Hc\cdot A_\perp}\right)+\overline{A}'\left(S_{\varepsilon H\cdot A_\perp},S_{\varepsilon c\cdot A_\perp}\right)\right\}\begin{pmatrix}I & -S_{HH\cdot\hat{\alpha}_\perp}^{-1}S_{Hc\cdot\hat{\alpha}_\perp}\\ 0 & 1\end{pmatrix}\\ &\quad\times\begin{pmatrix}S_{HH\cdot\hat{\alpha}_\perp} & 0\\ 0 & S_{cc\cdot H,\hat{\alpha}_\perp}\end{pmatrix}^{-1}\begin{pmatrix}I & 0\\ -S_{cH\cdot\hat{\alpha}_\perp}S_{HH\cdot\hat{\alpha}_\perp}^{-1} & 1\end{pmatrix}\\ &\quad\times\left\{\begin{pmatrix}S_{HH\cdot\hat{\alpha}_\perp}\\ S_{cH\cdot\hat{\alpha}_\perp}\end{pmatrix}\varphi\psi'+\begin{pmatrix}S_{H\varepsilon\cdot\hat{\alpha}_\perp}\\ S_{c\varepsilon\cdot\hat{\alpha}_\perp}\end{pmatrix}\overline{A}\right\}\overline{\psi}\left\{1+\mathrm{o}\left(T^{-\xi/2}\right)\right\}.\end{aligned}$$

and

$$\begin{aligned} 0 &\overset{a.s.}{=} \tilde{\psi}_{\perp}'\left(\psi\varphi' S_{HH\cdot A_{\perp}} + \overline{A}' S_{\varepsilon H\cdot A_{\perp}}\right) S_{HH\cdot\hat{\alpha}_{\perp}}^{-1}\left(S_{HH\cdot\hat{\alpha}_{\perp}}\varphi + S_{H\varepsilon\cdot\hat{\alpha}_{\perp}}\overline{A\psi}\right) \\ &\qquad \times\left\{1+\mathrm{o}\left(T^{-\xi/2}\right)\right\} \\ &\quad + \tilde{\psi}_{\perp}'\psi\varphi'\left(S_{Hc\cdot A_{\perp}} - S_{HH\cdot A_{\perp}} S_{HH\cdot\hat{\alpha}_{\perp}}^{-1} S_{Hc\cdot\hat{\alpha}_{\perp}}\right) \\ &\qquad \times S_{cc\cdot H,\hat{\alpha}_{\perp}}^{-1}\left(0 + S_{c\varepsilon\cdot H,\hat{\alpha}_{\perp}}\right)\overline{A\psi}\left\{1+\mathrm{o}\left(T^{-\xi/2}\right)\right\} \\ &\quad + \tilde{\psi}_{\perp}'\overline{A}'\left(S_{\varepsilon c\cdot A_{\perp}} - S_{\varepsilon H\cdot A_{\perp}} S_{HH\cdot\hat{\alpha}_{\perp}}^{-1} S_{Hc\cdot\hat{\alpha}_{\perp}}\right) \\ &\qquad \times S_{cc\cdot H,\hat{\alpha}_{\perp}}^{-1}\left(0 + S_{c\varepsilon\cdot H,\hat{\alpha}_{\perp}}\right)\overline{A\psi}\left\{1+\mathrm{o}\left(T^{-\xi/2}\right)\right\} \end{aligned} \tag{F.30}$$

Now, (F.18), (F.19) show

$$\tilde{\psi}_{\perp}'\left(M_{\psi}'\right)^{-1} = \tilde{\psi}_{\perp}'\left(\overline{\psi}_{\perp}, \psi S_{\beta\beta}^{1/2}\right) \overset{a.s.}{=} (I_{m-r}, 0) + \mathrm{o}(T^{-\xi/2}) \tag{F.31}$$

This in combination with Lemma A.4, (F.23) shows that

$$\tilde{\psi}_{\perp}' S_{AA\cdot A_{\perp}}\tilde{\psi}_{\perp} \overset{a.s.}{=} \psi_{\perp}'\Omega_{AA\cdot A_{\perp}}\psi_{\perp} + \mathrm{o}(T^{-\xi/2}), \tag{F.32}$$

$$\tilde{\psi}_{\perp}' S_{AH\cdot A_{\perp}} S_{HH\cdot A_{\perp}}^{-1} S_{HA\cdot A_{\perp}}\tilde{\psi}_{\perp} \overset{a.s.}{=} \mathrm{o}\left(T^{-\xi}\right), \tag{F.33}$$

while the concentrated model equation (F.27) and Lemmas A.1, A.5, A.6 give

$$\tilde{\psi}_{\perp}' S_{A\varepsilon\cdot A_{\perp}} \overset{a.s.}{=} \mathrm{O}(1). \tag{F.34}$$

Using (F.32)-(F.34) it is seen that

$$S_{HH\cdot\hat{\alpha}} \overset{a.s.}{=} S_{HH\cdot A_{\perp}}\left\{1+\mathrm{o}(T^{-\xi})\right\}, \tag{F.35}$$

$$S_{HH\cdot A_{\perp}}^{-1/2} S_{H\varepsilon\cdot\hat{\alpha}_{\perp}} \overset{a.s.}{=} S_{HH\cdot A_{\perp}}^{-1/2} S_{H\varepsilon\cdot A_{\perp}} + \mathrm{o}(T^{-\xi/2}). \tag{F.36}$$

and $S_{HH\cdot A_{\perp}}^{-1/2} S_{H\varepsilon\cdot\hat{\alpha}_{\perp}} \overset{a.s.}{=} S_{HH\cdot A_{\perp}}^{-1/2} S_{H\varepsilon\cdot A_{\perp}} + \mathrm{o}(T^{-\xi/2})$.

Further, (A.1), (A.4), (A.5) in Lemma A.5 imply that

$$S_{cc\cdot H,A_{\perp}} \overset{a.s.}{=} S_{cc\cdot H}\left\{1+\mathrm{o}\left(T^{-\xi}\right)\right\}, \tag{F.37}$$

(F.31), (F.32) imply

$$\tilde{\psi}_{\perp}' S_{AA\cdot H,A_{\perp}}\tilde{\psi}_{\perp} \overset{a.s.}{=} \tilde{\psi}_{\perp}' S_{AA\cdot A_{\perp}}\tilde{\psi}_{\perp} + \mathrm{o}\left(T^{-\xi}\right) \overset{a.s.}{=} \mathrm{O}(1), \tag{F.38}$$

(A.1), (A.4), (A.5) in Lemma A.5 imply

$$S_{\varepsilon c\cdot H,A_{\perp}} S_{cc\cdot H}^{-1/2} \overset{a.s.}{=} \mathrm{o}\left(T^{-\xi/2}\right), \tag{F.39}$$

(F.31) and (A.1), (A.4), (A.5) in Lemma A.5 imply

$$\tilde{\psi}'_{\perp} S_{Ac\cdot H,A_{\perp}} S_{cc\cdot H}^{-1/2} \overset{a.s.}{=} \mathrm{o}\left(T^{-\xi/2}\right), \tag{F.40}$$

(F.37), (F.38), (F.40) imply

$$S_{cc\cdot H,\hat{\alpha}_{\perp}} \overset{a.s.}{=} S_{cc\cdot H}\left\{1 + \mathrm{o}\left(T^{-\xi}\right)\right\} \tag{F.41}$$

(F.38), (F.40) and (A.1), (A.4), (A.5) in Lemma A.5 imply

$$S_{cc\cdot H}^{-1/2} S_{c\varepsilon\cdot H,\hat{\alpha}_{\perp}} \overset{a.s.}{=} S_{cc\cdot H}^{-1/2} S_{c\varepsilon\cdot H} \overset{a.s.}{=} \mathrm{o}(T^{-\xi/2}). \tag{F.42}$$

and by combining Lemmas C.3 and C.1

$$\liminf_{T\to\infty} S_{cc\cdot H} \overset{a.s.}{>} 0. \tag{F.43}$$

It then follows that with probability one (C.14) is equivalent to

$$\begin{aligned} 0 \overset{a.s.}{=} & \left\{\tilde{\psi}'_{\perp}\psi S_{\beta\beta} + \psi'_{\perp}\overline{A}'_{\omega} + \mathrm{o}\left(T^{-\xi}\right)\right\}\left\{1 + \mathrm{o}\left(T^{-\xi/2}\right)\right\} \\ & + \tilde{\psi}'_{\perp}\left\{\psi\varphi' + \mathrm{o}\left(T^{-\xi/2}\right) S_{HH}^{-1/2}\right\}\left(S_{Hc\cdot A_{\perp}} - S_{Hc\cdot\hat{\alpha}_{\perp}}\right)\mathrm{o}\left(T^{-\xi/2}\right) + \mathrm{o}\left(T^{-\xi}\right) \end{aligned} \tag{F.44}$$

where (C.18) in Lemma A.9 has been applied to the first term and (F.24), (F.35), (F.41), (F.39), (F.42), (F.43) have been used for the second term.

Finally, the term $S_{Hc\cdot A_{\perp}} - S_{Hc\cdot\hat{\alpha}_{\perp}}$ is studied. To this end note that for all $\eta > 0$

$$\begin{aligned} \tilde{\psi}'_{\perp} S_{Ac\cdot A_{\perp}} &= \tilde{\psi}'_{\perp}\left(\overline{A}' - S_{AA_{\perp}} S_{A_{\perp}A_{\perp}}^{-1} A'_{\perp}\right)\left(A\psi\varphi' S_{Hc} + S_{\varepsilon c}\right) \\ &= \tilde{\psi}'_{\perp}\psi S_{\beta\beta}^{1/2} S_{\beta\beta}^{-1/2} S_{\beta c} + \tilde{\psi}'_{\perp}\left(\overline{A}' - S_{AA_{\perp}} S_{A_{\perp}A_{\perp}}^{-1} A'_{\perp}\right) S_{\varepsilon c} \\ &= \mathrm{o}_{\mathcal{P}}\left(T^{-\xi/2+\eta}\right) \end{aligned} \tag{F.45}$$

since $S_{\beta\beta}^{-1/2} S_{\beta c} = \mathrm{o}_{\mathcal{P}}\left(T^{\eta}\right)$ by Lemma A.10 and this together with (F.18) show the first term is $\mathrm{o}_{\mathcal{P}}\left(T^{-\xi/2+\eta}\right)$ while Lemma A.2 together with (F.31) and (A.1) show the second term is $\mathrm{o}(T^{-\xi/2})$ with probability one. Therefore using (F.32), (F.33), (F.45) it follows that

$$\begin{aligned} S_{HH}^{-1/2}\left(S_{Hc\cdot A_{\perp}} - S_{Hc\cdot\hat{\alpha}_{\perp}}\right) &= S_{HH}^{-1/2} S_{HA\cdot A_{\perp}}\tilde{\psi}_{\perp}\left(\tilde{\psi}'_{\perp} S_{AA\cdot A_{\perp}}\tilde{\psi}'_{\perp} S_{Ac\cdot A_{\perp}}\right)^{-1}\tilde{\psi}'_{\perp} S_{Ac\cdot A_{\perp}} \\ &= \mathrm{o}_{\mathcal{P}}\left\{T^{-\xi+\eta}\right\}. \end{aligned}$$

Therefore using (F.18), (F.19) then (F.44) can be rewritten as

$$0 \overset{a.s.}{=} \left\{\tilde{\psi}'_{\perp}\psi S_{\beta\beta} + \psi'_{\perp}\overline{A}'_{\omega} + \mathrm{o}\left(T^{-\xi}\right)\right\}\left\{1 + \mathrm{o}\left(T^{-\xi/2}\right)\right\} + \mathrm{o}_{\mathcal{P}}\left\{T^{-2\xi+\eta}\right\} + \mathrm{o}\left(T^{-\xi}\right)$$

as desired. ∎

F.4.4 *Test for simple hypothesis on* $\alpha_\perp$ *when* $\beta = H\varphi$, $\alpha = A\psi$ *and* $\beta_c' \in \mathbf{R}^r$

This result generalises that found in Section B.3. The argument is simply extended drawing on (A.5) in Lemma A.7.

The following the likelihood ratio test for a simple hypothesis on $\alpha_\perp$ is considered

$$H_{\alpha_\perp^\circ}(r): \qquad \alpha_\perp = \left(A_\perp, \overline{A}\psi_\perp^\circ\right)$$

for some known matrix $\psi_\perp^\circ$.

The following result modifies Lemma B.1.

Lemma F.23 *Suppose equation* (2.6), $\alpha_\perp = (A_\perp, \overline{A}\psi_\perp^\circ)$, *the hypotheses* $H_\alpha^c(r)$, $H_\beta^c(r)$, *condition* (i) *and the Assumptions 2.12, 2.13 are satisfied. Then*

$$LR\left\{ H_{\alpha_\perp^\circ}^c(r) \middle| H_\alpha^c(r), H_\beta^c(r)\right\} = tr\left\{(\psi_\perp'\Omega_{AA\cdot A_\perp}\psi_\perp)^{-1} T\psi_\perp'\overline{A}_\omega' S_{\varepsilon\beta}S_{\beta\beta}^{-1}S_{\beta\varepsilon}\overline{A}_\omega\psi_\perp\right\} + o_{\mathcal{P}}(1).$$

Proof of Lemma F.23. As in the proof of Lemma B.1

$$\begin{aligned} LR\left\{ H_{\alpha_\perp^\circ}(r) \middle| H_\alpha(r), H_\beta(r)\right\} &= T\left\{\log\frac{|\psi_\perp^{\circ\prime}S_{AA\cdot A_\perp}\psi_\perp^\circ|}{|\psi_\perp^{\circ\prime}S_{AA\cdot H,c,A_\perp}\psi_\perp^\circ|} - \log\frac{\left|\hat{\psi}_\perp' S_{AA\cdot A_\perp}\hat{\psi}_\perp\right|}{\left|\hat{\psi}_\perp' S_{AA\cdot H,c,A_\perp}\hat{\psi}_\perp\right|}\right\} \\ &= \operatorname{tr}(TD_1^c) - \operatorname{tr}(TD_2^c) + \mathrm{O}\left(\left\|\tilde{\psi}_\perp - \psi_\perp\right\|^3\right) \end{aligned}$$

where $\tilde{\psi}_\perp - \psi_\perp \overset{a.s.}{\to} 0$ by the same argument as that of (A.8) in Lemma A.9 and

$$\begin{aligned} D_1^c &= (\psi_\perp' S_{AA\cdot A_\perp}\psi_\perp)^{-1}\left(\tilde{\psi}_\perp - \psi_\perp\right)' S_{AA\cdot A_\perp,\overline{A}\psi_\perp}\left(\tilde{\psi}_\perp - \psi_\perp\right), \\ D_2^c &= (\psi_\perp' S_{AA\cdot H,c,A_\perp}\psi_\perp)^{-1}\left(\tilde{\psi}_\perp - \psi_\perp\right)' S_{AA\cdot H,c,A_\perp,\overline{A}\psi_\perp}\left(\tilde{\psi}_\perp - \psi_\perp\right). \end{aligned}$$

Using Lemma A.2 and (A.4) of Lemma A.7 it is seen that $\psi_\perp' S_{AA\cdot A_\perp}\psi_\perp \overset{a.s.}{=} \psi_\perp'\Omega_{AA\cdot A_\perp}\psi_\perp + \mathrm{o}(T^{-\xi/2})$ whereas $\psi_\perp' S_{AA\cdot H,c,A_\perp}\psi_\perp \overset{a.s.}{=} \psi_\perp'\Omega_{AA\cdot A_\perp}\psi_\perp + \mathrm{o}(T^{-\xi/2})$. Thus applying the definition $\tilde{\psi}_\perp - \psi_\perp = \overline{\psi}_\perp U_T$ it follows

$$D_1^c - D_2^c \overset{a.s.}{=} (\psi_\perp'\Omega_{AA\cdot A_\perp}\psi_\perp)^{-1} U_T'\overline{\psi}'\left(S_{AA\cdot\alpha_\perp} - S_{AA\cdot H,c,\alpha_\perp,}\right)\overline{\psi}U_T\left\{1 + \mathrm{o}\left(T^{-\xi}\right)\right\}.$$

Equations (A.4), (A.5) of Lemma A.7 show

$$S_{cc\cdot H,\alpha_\perp} \overset{a.s.}{=} S_{cc\cdot H}\left\{1 + \mathrm{o}\left(T^{-\xi}\right)\right\}$$

while (A.1), (A.5) of Lemma A.7 show

$$M_\psi' S_{ac\cdot H,\alpha_\perp} S_{cc\cdot H}^{-1/2} \overset{a.s.}{=} \mathrm{o}\left(T^{-\xi/2}\right).$$

It then follows from (F.18) in Lemma F.22 that

$$U_T'\overline{\psi}'S_{aa\cdot H,c,\alpha_\perp}\overline{\psi}U_T \overset{a.s.}{=} U_T'\overline{\psi}'S_{aa\cdot H,\alpha_\perp}\overline{\psi}U_T + \mathrm{o}\left(T^{-\xi}\right)$$

and therefore

$$D_1^c - D_2^c \overset{a.s.}{=} \left(\psi_\perp'\Omega_{AA\cdot A_\perp}\psi_\perp\right)^{-1} U_T'\overline{\psi}'\left(S_{AA\cdot\alpha_\perp} - S_{AA\cdot H,\alpha_\perp,}\right)\overline{\psi}U_T\left\{1+\mathrm{o}\left(T^{-\xi}\right)\right\} + \mathrm{o}\left(T^{-\xi}\right).$$

The remainder of the proof of Lemma B.1 can then be followed using (F.20) in Lemma F.22 rather than (A.9) in Lemma A.9. ■

F.4.5 ***Test for*** $\alpha = A\psi$ ***when*** $\beta = H\varphi,\ \beta_c' \in \mathbf{R}^r$

This result generalises Theorem 2.8.

Corollary F.24 *Suppose equation* (2.6), *the hypothesis* $H_\alpha^c(r)$, *the conditions* (i), (iv) *and the Assumptions 2.12, 2.13 are satisfied. Then*

$$LR\left\{ H_\alpha^c\left(r\right) \middle| H_\beta^c(r)\right\} \xrightarrow{\mathcal{D}} \chi^2\left\{r\left(m-r\right)\right\}$$

Proof of Lemma F.24. Follow the proof of Theorem 2.8 by combining the result in (2.3) in Theorem 2.2 with Lemmas A.5, F.23 ■

F.4.6 *Test for* $\beta_c = 0$ *when* $\alpha_\perp$ *is known and* $\beta = H\varphi$

This result generalises that found in Section B.5. In particular the concentrated model equation is

$$\overline{\alpha}' R_{0,t} = \varphi' H' R_{1,t} + \beta_c' (1|Z_{t-1}) + \omega \alpha_\perp' R_{0,t} + \overline{\alpha}_\omega' \hat{\varepsilon}_t \tag{F.46}$$

Thus the likelihood ratio test statistic for $\beta_c = 0$ is given by

$$LR\left\{ H(r) | H^c_{\alpha_\perp^\circ}(r) \right\} = -T \log\det\left(I_r - S^{-1}_{\alpha\alpha\cdot H,\alpha_\perp} S_{\alpha c\cdot H,\alpha_\perp} S^{-1}_{cc\cdot H,\alpha_\perp} S_{c\alpha\cdot H,\alpha_\perp} \right).$$

For the asymptotic distribution it is necessary that condition (ii) is satisfied. The result is discussed in two Lemmas. First, the test statistic is rewritten without using condition (ii) and next the asymptotic result is given

Lemma F.25 *Suppose equation* (2.6), $\alpha_\perp = \alpha_\perp^\circ$, *the hypothesis* $H(r)$, *condition* (i) *and Assumption 2.12 are satisfied. Then*

$$LR\left\{ H(r) | H^c_{\alpha_\perp^\circ}(r) \right\} \overset{a.s.}{=} \operatorname{tr}\left(\Omega^{-1}_{\alpha\alpha\cdot\alpha_\perp} T \overline{\alpha}_\omega' S_{\varepsilon c\cdot H} S^{-1}_{cc\cdot H} S_{c\varepsilon\cdot H} \overline{\alpha}_\omega \right) + \mathrm{o}(1).$$

Proof of Lemma F.25. *First,* as in (B.7) in Lemma B.2

$$S_{\beta\beta}^{-1/2} S_{\alpha\alpha\cdot H,\alpha_\perp} S_{\beta\beta}^{-1/2} \overset{a.s.}{=} S_{\beta\beta}^{-1/2} \Omega_{\alpha\alpha\cdot\alpha_\perp} S_{\beta\beta}^{-1/2} + \mathrm{o}(1).$$

Secondly, by Lemma A.5 and (A.1) of Lemma A.7

$$S_{cc\cdot H}^{-1/2} S_{cc\cdot H,\alpha_\perp} S_{cc\cdot H}^{-1/2} = 1 - S_{cc\cdot H}^{-1/2} S_{c\alpha_\perp\cdot H} S^{-1}_{\alpha_\perp\alpha_\perp\cdot H} S_{\alpha_\perp c\cdot H} S_{cc\cdot H}^{-1/2} \overset{a.s.}{=} 1 + \mathrm{o}(1). \tag{F.47}$$

Thirdly, Since it is tested that $\beta_c = 0$ the partial model equation (F.46) shows $S_{\alpha c\cdot H,\alpha_\perp} = \overline{\alpha}_\omega' S_{\varepsilon c\cdot H,\alpha_\perp}$. Further,

$$\overline{\alpha}_\omega' \left(S_{\varepsilon c\cdot H,\alpha_\perp} - S_{\varepsilon c\cdot H} \right) S_{cc\cdot H}^{-1/2} = \overline{\alpha}_\omega' S_{\varepsilon\varepsilon\cdot H} \alpha_\perp S^{-1}_{\alpha_\perp\alpha_\perp\cdot H} \alpha_\perp' S_{\varepsilon c\cdot H} S_{cc\cdot H}^{-1/2} \overset{a.s.}{=} \mathrm{o}(1),$$

by Lemmas A.5, A.6 and the result follows. ■

F.4.7 *Test for* $\beta_c = 0$.

Corollary F.26 *Suppose equation* (2.6), *the hypothesis* $H(r)$, *the conditions* (i), (ii) *and the Assumptions 2.12, 2.13 are satisfied. Then*

$$LR\left\{ H(r) \middle| H^c(r) \right\} \xrightarrow{\mathcal{D}} \chi^2(r).$$

Proof of Corollary F.26. As in the proof of Theorem 2.7 let $L(\alpha_\perp, \beta, \beta_c, \Omega)$ denote the likelihood concentrated with respect to the lagged differences. The likelihood ratio test statistic is then

$$Q\left\{ H(r) \middle| H_c(r) \right\} = \max_{\alpha_\perp,\beta,\Omega} L(\alpha_\perp, \beta, 0, \Omega) \Big/ \max_{\alpha_\perp,\beta,\beta_c,\Omega} L(\alpha_\perp, \beta, \beta_c, \Omega).$$

Extend this fraction by the maximised likelihood when $\alpha_\perp$ is known to see

$$\begin{aligned} & Q\left\{ H(r) \middle| H_c(r) \right\} \\ = & \left(\frac{\max_{\alpha_\perp,\beta,\Omega} L(\alpha_\perp, \beta, 0, \Omega)}{\max_{\beta,\Omega} L(\alpha_\perp^\circ, \beta, 0, \Omega)} \right) \left(\frac{\max_{\beta,\Omega} L(\alpha_\perp^\circ, \beta, 0, \Omega)}{\max_{\beta,\beta_c,\Omega} L(\alpha_\perp^\circ, \beta, \beta_c, \Omega)} \right) \left(\frac{\max_{\beta,\beta_c,\Omega} L(\alpha_\perp^\circ, \beta, \beta_c, \Omega)}{\max_{\alpha_\perp,\beta,\beta_c,\Omega} L(\alpha_\perp, \beta, \beta_c, \Omega)} \right). \end{aligned}$$

The Lemmas B.1, F.23 show that asymptotically the tests for simple hypotheses on $\alpha_\perp$ depend on the same statistic, which does not depend on β_c and therefore the first and the last term cancel. Using the Lemma F.25 it is therefore seen that

$$LR\left\{ H(r) \middle| H_c(r) \right\} = \operatorname{tr}\left(\Omega_{\alpha\alpha\cdot\alpha_\perp}^{-1} T \overline{\alpha}_\omega' S_{\varepsilon c\cdot H} S_{cc\cdot H}^{-1} S_{c\varepsilon\cdot H} \overline{\alpha}_\omega \right) + o_{\mathcal{P}}(1).$$

Under condition (ii) the Lemma A.10 shows that for some $\eta > 0$

$$S_{cH} S_{HH}^{-1} S_{Hc} = o_{\mathcal{P}}\left(T^{\eta-1} \right),$$

and the Lemmas A.5, A.6 then show that

$$\begin{aligned} S_{cc\cdot H} &= 1 + o_{\mathcal{P}}(1), \\ T^{1/2} S_{\varepsilon c\cdot H} &= T^{1/2} S_{\varepsilon c} + o_{\mathcal{P}}\left(T^{\eta-\xi} \right) = T^{-1/2} \sum_{t=1}^{T} \varepsilon_t + o_{\mathcal{P}}\left(T^{\eta-\xi} \right). \end{aligned}$$

Since η can be chosen $\eta < \xi$ it follows that

$$LR\left\{ H(r) \middle| H_c(r) \right\} = \operatorname{tr}\left\{ \Omega_{\alpha\alpha\cdot\alpha_\perp}^{-1} \left(T^{-1/2} \overline{\alpha}_\omega' \sum_{t=1}^{T} \varepsilon_t \right)^{\otimes 2} \right\} + o_{\mathcal{P}}(1).$$

The desired result then follows from Brown's (1971) martingale Central Limit Theorem, see the proof of Lemma A.13. ■

F.4.8 *Test for* $\beta = H\varphi$ *when* $\alpha_\perp$ *is known and* $\beta_c' \in \mathbf{R}^r$

Following the proof of Lemma B.2 in Section B.5 the partial model equation is

$$\overline{\alpha}' R_{0,t} = \beta' R_{1,t} + \beta_c' (1|Z_{t-1}) + \omega\alpha_\perp' R_{0,t} + \overline{\alpha}_\omega' \hat{\varepsilon}_t$$

where $\omega = \Omega_{\alpha\alpha_\perp}\Omega^{-1}_{\alpha_\perp\alpha_\perp}$. The likelihood ratio test statistic for hypothesis H_β is therefore

$$LR\left\{ H_\beta^c(r) \middle| H^c_{\alpha_\perp^\circ}(r)\right\} = -T\log\det\left(I_r - S^{-1}_{\alpha\alpha\cdot H,c,\alpha_\perp} S_{\alpha H_\perp\cdot H,c,\alpha_\perp} S^{-1}_{H_\perp H_\perp\cdot H,c,\alpha_\perp} S_{H_\perp\alpha\cdot H,c,\alpha_\perp}\right).$$

The following result modifies Lemma B.2

Lemma F.27 *Suppose equation* (2.6), *the hypothesis* $H_\beta^c(r)$, *condition* (i), *and Assumption 2.12 are satisfied. Then*

$$LR\left\{ H_\beta^c(r) \middle| H^c_{\alpha_\perp^\circ}(r)\right\} \overset{\mathcal{D}}{\to} \chi^2\left\{r(p-s)\right\}$$

Proof of Lemma F.27. *First,* from Lemma A.2 and (A.4) of Lemma A.5 it follows that

$$S_{cc\cdot H,\alpha_\perp} \overset{a.s.}{=} S_{cc\cdot H}\left\{1 + \mathrm{o}\left(T^{-\xi}\right)\right\},$$

which together with Lemma A.4 and (A.1), (A.5) of Lemma A.5 imply

$$S^{-1/2}_{\beta\beta} S_{\alpha c\cdot H,\alpha_\perp} S^{-1/2}_{cc\cdot H} \overset{a.s.}{=} \mathrm{o}\left(T^{-\xi/2}\right).$$

This again, together with (A.1), (A.4) shows that

$$S^{-1/2}_{\beta\beta} S_{\alpha\alpha\cdot H,c,\alpha_\perp} S^{-1/2}_{\beta\beta} \overset{a.s.}{=} S^{-1/2}_{\beta\beta}\Omega_{\alpha\alpha\cdot\alpha_\perp} S^{-1/2}_{\beta\beta} + \mathrm{o}(1)$$

Secondly, it is argued that

$$S_{H_\perp H_\perp\cdot H,c,\alpha_\perp} \overset{a.s.}{=} S_{H_\perp H_\perp\cdot H,c}\left\{1 + \mathrm{o}(1)\right\},$$

since $S^{-1/2}_{H_\perp H_\perp\cdot H,c} S_{H_\perp\alpha_\perp\cdot H,c} \overset{a.s.}{\to} 0$ by Lemma A.5 and $S_{\alpha_\perp\alpha_\perp\cdot H,c} \overset{a.s.}{\to} \Omega_{\alpha_\perp\alpha_\perp}$ by Lemma A.5 and (A.1) of Lemma A.7.

Thirdly, by the partial model equation $S_{\alpha H_\perp\cdot H,c,\alpha_\perp} = \overline{\alpha}_\omega' S_{\varepsilon H_\perp\cdot H,c,\alpha_\perp}$ and therefore

$$\begin{aligned} & \overline{\alpha}_\omega'\left(S_{\varepsilon H_\perp\cdot H,c,\alpha_\perp} - S_{\varepsilon H_\perp\cdot H,c}\right) S^{-1/2}_{H_\perp H_\perp\cdot H,c} \\ = \; & \overline{\alpha}_\omega'\left\{S_{\varepsilon\varepsilon} - S_{\varepsilon(H,c)} S^{-1}_{(H,c)(H,c)} S_{(H,c)\varepsilon}\right\}\alpha_\perp S^{-1}_{\alpha_\perp\alpha_\perp\cdot H,c}\alpha_\perp' S_{\varepsilon H_\perp\cdot H,c} S^{-1/2}_{H_\perp H_\perp\cdot H,c} \overset{a.s.}{=} \mathrm{o}(1) \end{aligned}$$

by Lemmas A.1, A.5 and (A.1) of Lemma A.5. In particular by Lemma A.4

$$S^{-1/2}_{\beta\beta} S_{\alpha H_\perp\cdot H,c,\alpha_\perp} S^{-1/2}_{H_\perp H_\perp\cdot H,c} \overset{a.s.}{=} S^{-1/2}_{\beta\beta}\overline{\alpha}_\omega' S_{\varepsilon H_\perp\cdot H,c} S^{-1/2}_{H_\perp H_\perp\cdot H,c} + \mathrm{o}(1).$$

Finally, in combination

$$LR\left\{ H_\beta^c(r) \middle| H^c_{\alpha_\perp^\circ}(r)\right\} \overset{a.s.}{=} -T\log\det\left(I_p - \Omega^{-1}_{\alpha\alpha\cdot\alpha_\perp}\overline{\alpha}_\omega' S_{\varepsilon H_\perp\cdot H,c} S^{-1}_{H_\perp H_\perp\cdot H,c} S_{H_\perp\varepsilon\cdot H,c}\overline{\alpha}_\omega\right) + \mathrm{o}(1).$$

the result follows by extending the Mixed Gaussian Central Limit Theorem A.14 to include regression on a constant. ∎

F.4.9 *Test for* $\beta = H\varphi$

The following result is a modified version of Theorem 2.7.

Corollary F.28 *Suppose equation* (2.6), *the hypothesis* $H(r)$, *condition* (i), *and the Assumptions 2.12, 2.13 are satisfied. Then*

$$LR\left\{ H_{\beta}^{c}(r) \middle| H_{c}(r)\right\} \overset{\mathcal{D}}{\to} \chi^{2}\left\{r(p-s)\right\}.$$

Proof of Corollary F.28. Follow the proof of Corollary F.26 using Lemmas F.23, F.27. ■

F.5 *Counter Examples*

Consider the probability measure given by the equation (2.5), that is

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

where ε_t has a standard normal distribution. Suppose $X_0 = 0$. Define for $j = 1, 2$

$$\varepsilon_j = \varepsilon_t^{(j)}, \qquad S_j = \sum_{s=1}^{t} \varepsilon_s^{(j)}, \qquad SS_j = \sum_{s=1}^{t} \sum_{r=1}^{s-1} \varepsilon_r^{(j)}.$$

Define univariate standard Brownian motions V, W and let

$$\widetilde{W} = \int_0^1 W_u du, \qquad \widetilde{\widetilde{W}} = \int_0^1 \int_0^u W_t dt du, \qquad \int W^2 = \int_0^1 W_u^2 du,$$

$$D = \int \widetilde{W}^2 \int W^2 - \left(\int \widetilde{W} W \right)^2$$

Then

$$\Delta X_t = \begin{pmatrix} S_2 + \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}, \qquad X_t = \begin{pmatrix} SS_2 + S_2 \\ S_2 \end{pmatrix}$$

and

$$\begin{aligned}
\begin{pmatrix} T^{-1/2} & \\ & 1 \end{pmatrix} S_{00} \begin{pmatrix} T^{-1/2} & \\ & 1 \end{pmatrix} &\xrightarrow{\mathcal{D}} \begin{pmatrix} \int W^2 & 0 \\ 0 & 1 \end{pmatrix}, \\
\begin{pmatrix} T^{-1} & \\ & 1 \end{pmatrix} S_{10} \begin{pmatrix} T^{-1} & \\ & 1 \end{pmatrix} &\xrightarrow{\mathcal{D}} \begin{pmatrix} \int \widetilde{W} W & \int \widetilde{W} dW \\ \int W^2 & \int W dW \end{pmatrix}, \\
T^{-1} \begin{pmatrix} T^{-1} & \\ & 1 \end{pmatrix} S_{10} \begin{pmatrix} T^{-1} & \\ & 1 \end{pmatrix} &\xrightarrow{\mathcal{D}} \begin{pmatrix} \int \widetilde{W}^2 & \int \widetilde{W} W \\ \int \widetilde{W} W & \int W^2 \end{pmatrix}, \\
T^{1/2} \begin{pmatrix} T^{-1} & \\ & 1 \end{pmatrix} S_{0c} &\xrightarrow{\mathcal{D}} \begin{pmatrix} \widetilde{W}_1 \\ W_1 \end{pmatrix}, \\
T^{-1/2} \begin{pmatrix} T^{-1} & \\ & 1 \end{pmatrix} S_{1c} &\xrightarrow{\mathcal{D}} \begin{pmatrix} \widetilde{\widetilde{W}}_1 \\ \widetilde{W}_1 \end{pmatrix}, \\
S_{cc} &= 1, \\
T^{1/2} S_{\varepsilon c} &\xrightarrow{\mathcal{D}} \begin{pmatrix} V_1 \\ W_1 \end{pmatrix}, \\
\begin{pmatrix} T^{-1/2} & \\ & 1 \end{pmatrix} S_{1\varepsilon} &\xrightarrow{\mathcal{D}} \begin{pmatrix} \int \widetilde{W} dV & \int \widetilde{W} dW \\ \int W dV & \int W dW \end{pmatrix}
\end{aligned}$$

F.5.1 ***Necessity of assumption (c) for*** $LR\{H_\alpha(r)|H_\beta(r)\}$

The expression in Lemma B.1 gives

$$LR\left\{H_{\alpha_\perp^\circ}(r)\,|H(r)\right\} \overset{a.s.}{=} \sum \varepsilon_s X_2 \left(\sum X_2^2\right)^{-1} \sum X_2 \varepsilon_2 \approx \frac{\left(\int W dW\right)^2}{\int W^2}.$$

F.5.2 ***Necessity of*** (ii) ***when testing*** $\beta_c = 0$

The expression in Lemma F.25 gives

$$LR \overset{a.s.}{=} -T \log\det\left(I_r - \overline{\alpha}_\omega' S_{\varepsilon c\cdot 1} S_{cc\cdot 1}^{-1} S_{c\varepsilon\cdot 1} \overline{\alpha}_\omega\right) + \mathrm{o}(1).$$

Below it will be argued that

$$\begin{aligned} T^{1/2}\overline{\alpha}_\omega' S_{\varepsilon c\cdot 1} &\overset{\mathcal{D}}{\to} \int F dV - \int W dW \int F dW, \\ S_{cc\cdot 1} &\overset{\mathcal{D}}{\to} \int F^2 du. \end{aligned}$$

where

$$F = 1 - \widetilde{\widetilde{W}}_1 \frac{\left(\widetilde{W}\middle| W\right)}{\int \left(\widetilde{W}\middle| W\right)^2} - \widetilde{W}_1 \frac{\left(W|\,\widetilde{W}\right)}{\int \left(W|\,\widetilde{W}\right)^2}.$$

It then follows that LR has a non-standard distribution

$$LR \overset{\mathcal{D}}{\to} \left(\frac{\int F dV}{\sqrt{\int F^2 du}} - \int W dW \frac{\int F dW}{\sqrt{\int F^2 du}}\right)^2$$

Now,

$$\begin{aligned} S_{cc\cdot 1} &= S_{cc} - S_{c1} S_{11}^{-1} S_{1c} \overset{\mathcal{D}}{\to} 1 - \left(\widetilde{\widetilde{W}}_1, \widetilde{W}_1\right) \frac{1}{D} \begin{pmatrix} \int W^2 & -\int \widetilde{W} W \\ -\int \widetilde{W} W & \int \widetilde{W}^2 \end{pmatrix} \begin{pmatrix} \widetilde{\widetilde{W}}_1 \\ \widetilde{W}_1 \end{pmatrix} \\ &= 1 - \frac{\widetilde{\widetilde{W}}_1^2}{\int \left(\widetilde{W}\middle| W\right)^2} - \frac{\widetilde{W}_1^2}{\int \left(W|\,\widetilde{W}\right)^2} + \frac{2\widetilde{W}_1 \widetilde{\widetilde{W}}_1 \int \widetilde{W} W}{D} = \int F^2 \end{aligned}$$

This is because

$$\begin{aligned} \int \left(\widetilde{W}\middle| W\right)\left(W|\,\widetilde{W}\right) &= \int \widetilde{W}\left(W|\,\widetilde{W}\right) - \frac{\int \widetilde{W} W}{\int W^2} \int W \left(W|\,\widetilde{W}\right) \\ &= -\int \widetilde{W} W \left\{1 - \frac{\left(\int \widetilde{W} W\right)^2}{\int W^2 \int \widetilde{W}^2}\right\} \end{aligned}$$

so that

$$\frac{\widetilde{W}_1 \widetilde{\widetilde{W}}_1 \int \left(\widetilde{W}\middle| W\right)\left(W|\widetilde{W}\right)}{\int \left(\widetilde{W}\middle| W\right)^2 \int \left(W|\widetilde{W}\right)^2} = \frac{-\widetilde{W}_1 \widetilde{\widetilde{W}}_1 \int \widetilde{W} W}{D},$$

and further

$$\begin{aligned} \widetilde{\widetilde{W}}_1 \int \left(\widetilde{W}\middle| W\right) &= \widetilde{\widetilde{W}}_1 \left(\widetilde{\widetilde{W}}_1 - \frac{\int \widetilde{W} W}{\int W^2} \widetilde{W}_1\right) = \widetilde{\widetilde{W}}_1^2 - \frac{\int \widetilde{W} W}{\int W^2} \widetilde{W}_1 \widetilde{\widetilde{W}}_1, \\ \widetilde{W}_1 \int \left(W|\widetilde{W}\right) &= \widetilde{W}_1 \left(\widetilde{W}_1 - \frac{\int \widetilde{W} W}{\int \widetilde{W}^2} \widetilde{\widetilde{W}}_1\right) = \widetilde{W}_1^2 - \frac{\int \widetilde{W} W}{\int \widetilde{W}^2} \widetilde{W}_1 \widetilde{\widetilde{W}}_1 \end{aligned}$$

Now,

$$\begin{aligned} T^{1/2}(1,0) S_{\varepsilon c \cdot 1} = & \quad T^{1/2}(1,0)\left(S_{\varepsilon c} - S_{\varepsilon 1} S_{11}^{-1} S_{1c}\right) \\ \stackrel{\mathcal{D}}{\rightarrow} & \quad V_1 - \frac{1}{D} \int \left(\widetilde{W}, W\right) dV \begin{pmatrix} \int W^2 & -\int \widetilde{W} W \\ -\int \widetilde{W} W & \int \widetilde{W}^2 \end{pmatrix} \begin{pmatrix} \widetilde{\widetilde{W}}_1 \\ \widetilde{W}_1 \end{pmatrix} \\ = & \quad \int dV - \frac{1}{D}\left\{ \widetilde{\widetilde{W}}_1 \left(\int \widetilde{W} dV \int W^2 - \int \widetilde{W} W \int W dV \right) \right. \\ & \qquad \left. + \widetilde{W}_1 \left(\int W dV \int \widetilde{W}^2 - \int \widetilde{W} W \int \widetilde{W} dV \right) \right\} \\ = & \quad \int F dV. \end{aligned}$$

Similarly

$$T^{1/2}(0,1) S_{\varepsilon c \cdot 1} \stackrel{\mathcal{D}}{\rightarrow} \int F dW.$$

Finally,

$$\overline{\alpha}'_\omega = \alpha - \frac{S_{\alpha \alpha_\perp}}{S_{\alpha_\perp \alpha_\perp}} \alpha'_\perp \stackrel{\mathcal{D}}{\rightarrow} \begin{pmatrix} 1 \\ 0 \end{pmatrix} - \int W dW \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$