Test for cointegration rank in general vector autoregressions

Bent Nielsen
Department of Economics, University of Oxford

22 September 2009

Johansen derived the asymptotic theory for his cointegration rank test statisic for a vector autoregression where the parameters are restricted so the process is integrated of order one. It is investigated to what extent these parameter restrictions are binding. The eigenvalues of Johansen's eigenvalue problem are shown to have the same consistency rates accross the parameter space. The test statistic is shown to have the usual asymptotic distribution as long as the possibilities of additional unit roots and of singular explosiveness are ruled out. To prove the results the convergence of stochastic integrals with respect to singular explosive processes is considered.

1 Introduction

The cointegration rank test statistic of Johansen (1988, 1995a) is analysed. This is a likelihood ratio test statistic in a vector autoregression. In the initial distributional analysis attention was restricted to the $\mathsf{I}(1)$-case thereby imposing restrictions on the parameter space of the vector autoregressive model. Subsequent research has shown that the same asymptotic distribution can arise in situations where these assumptions are not satisfied. Johansen and Schaumburg (1998) have shown this is the case for seasonally integrated processes while Nielsen (2001, 2005) has considered some scenarios involving explosive roots. In contrast to those results Johansen (1995b) shows that different asymptotic distributions arise in $\mathsf{I}(2)$-cases. In this paper results are given for the entire vector autoregressive parameter space.

Two types of results are given. First, the canonical correlations appearing in Johansen's eigenvalue problem are shown to be consistent in the entire parameter space. That is, the largest canonical correlations are shown to have positive limits, while the smallest canonical correlations vanish at a rate of T^{-1}. An almost sure version is given under some parameter restrictions.

Secondly, the parameter values are identified for which the rank test statistic has the usual asymptotic distribution. This happens quite generally in the parameter space with two exceptions. The first is that additional unit roots appearing in for instance $\mathsf{I}(2)$-case alter the asymptotic distribution. The second is that while regular explosive components are allowed the possibility of singular explosive components is ruled out. Such singular explosive components were noted by Anderson (1959) and have been discussed by Duflo, Senoussi, and Touati (1991), Phillips and Magdalinos

(2008) and Nielsen (2008). In the cointegration literature the main variants of the vector autoregressive model involve constants, linear trends and seasonal dummies. The presented asymptotic results cover these variants.

To establish these results the convergence of stochastic integrals with respect to singular explosive processes needs to be considered. The difficulty is that although singular explosive processes satisfy a Functional Central Limit Theorem they are not adapted to the natural filtration of the problem. This problem has been encountered previously in the context of integration with respect to mixing processes by de Jong and Davidson (2000). The solution considered here has more general integrands including various functions of random walks while the integrand is a singular explosive process which is a particular nice version of a mixingale.

Related results have been established previously for some mis-specification tests. Before conducting a rank test an investigator will be interested in checking the specification of the vector autoregression. Just as for the rank test asymptotic invariance with respect to the vector autoregressive parameters would be of interest. This has been established for lag length determination procedures by Nielsen (2006a, 2008), whereas the correlograms based on the Yule-Walker equations are not invariant, see Nielsen (2006b). Likewise Engler and Nielsen (2009) have shown that the empirical process of the residuals has the desired invariance properties as long as singular explosive roots are ruled out.

The paper is organsied so that §2 introduces the cointegration model. Granger-Johansen representations are given in §3. The asymptotic results are presented in §4. The convergence of stochastic integrals with respect to singular explosive processes is discussed in §5. Proofs are given in an appendix.

The following notation is used throughout the paper: For a matrix α let $\alpha^{\otimes 2} = \alpha\alpha'$. When α has full column rank then $\overline{\alpha} = \alpha(\alpha'\alpha)^{-1}$ whereas $\alpha_\perp$ is the orthogonal complement so $\alpha_\perp'\alpha = 0$ and $(\alpha, \alpha_\perp)$ is invertible. When α is symmetric then $\lambda_{\min}(\alpha)$ and $\lambda_{\max}(\alpha)$ are the smallest and the largest eigenvalue respectively. For matrices $||\alpha|| = \{\lambda_{\max}(\alpha^{\otimes 2})\}^{1/2}$ is the spectral norm, implying that $||\alpha^{-1}|| = \{\lambda_{\min}(\alpha^{\otimes 2})\}^{-1/2}$. While $\mathsf{E}(\varepsilon_t|\mathcal{F}_{t-1})$ is a conditional expectation the residuals of the least squares regression of Y_t on Z_t are denoted $(Y_t|Z_t) = Y_t - \sum_{s=1}^T Y_s Z_s'(\sum_{s=1}^T Z_s^{\otimes 2})^{-1} Z_t$ for a time series and $(y_u|z_u) = y_u - \int_0^1 y_v z_v' dv (\int_0^1 z_v z_v')^{-1} z_u$ for continuous processes.

2 Model and rank hypothesis

Suppose a p-dimensional time series, $X_{1-k}, \ldots, X_0, \ldots, X_T$ is available. The statistical model is then given by the vector autoregression

$$\mathsf{M}: \quad \Delta X_t = (\Pi, \Pi_d) \begin{pmatrix} X_{t-1} \\ d_{t-1} \end{pmatrix} + \sum_{j=1}^{k-1} \Gamma_j \Delta X_{t-j} + \mu D_{t-1} + \varepsilon_t, \tag{2.1}$$

for $t = 1, \ldots, T$, where the innovations ε_t are independently $\mathsf{N}(0, \Omega)$-distributed conditionally on the initial values $X_{1-k}, \ldots, X_0$ while d_t, D_{t-1} are deterministic components which are discussed below. Note, that in some cases the term $\Pi_d d_{t-1}$ is left out of the model equation (2.1). The normality assumption is necessary for defining a likelihood function. For the subsequent asymptotic analysis it can, however, be replaced by a martingale difference assumption. The parameters of the model are unrestricted so $\Pi, \Gamma_1, \ldots, \Gamma_{k-1}, \Omega \in \mathbf{R}^{p \times p}$, $\Pi_d, \mu \in \mathbf{R}^p$ and vary freely so Ω is positive definite. The likelihood function is defined accross this parameter space hence the interest in a distributional analysis of test statistics accross the parameter space.

Two types of deterministic terms are included. Let $\mu D_{t-1} = \mu_1 D_{1,t-1} + \mu_{\backslash 1} D_{\backslash 1,t-1}$, where $(d_t, D_{1,t})$ are polynomials like a constant, a linear trend, while $D_{\backslash 1,t}$ covers seasonal components. More formally,

$$d_t = d_{t-1} + (1, 0) D_{1,t-1}, \quad D_{1,t} = \mathbf{D}_1 D_{1,t-1}, \quad D_{\backslash 1,t} = \mathbf{D}_{\backslash 1} D_{\backslash 1,t-1}, \tag{2.2}$$

where $\mathbf{D}_1$ is a Jordan block of the form

$$\mathbf{D}_1 = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 1 \\ 0 & \cdots & 0 & 1 \end{pmatrix}, \tag{2.3}$$

while $\mathbf{D}_{\backslash 1}$ has eigenvalues on the complex unit circle except at one. Thus, $D_{\backslash 1,t}$ can include demeaned seasonal dummies with the property that they sum to zero. An example would be the biannual dummy $D_{\backslash 1,t} = (-1)^t$; see also the discussion of Johansen (1995, §5.8). In combination, $D_t = (D_{1,t}, D_{\backslash 1,t})$ satisfies the autoregressive equation $D_t = \mathbf{D} D_{t-1}$, where $\mathbf{D}$ is the blockdiagonal matrix $\mathbf{D} = \operatorname{diag}(\mathbf{D}_1, \mathbf{D}_{\backslash 1})$. It will be required that deterministic process satisfies $\operatorname{rank}(D_1, \ldots, D_{\dim \mathbf{D}}) = \dim \mathbf{D}$.

Johansen (1995) introduced five variants of deterministic terms. These are:

M_{lq}:	$d_t = t^2$, but omitted in regression,	$D_{1,t} = (t, 1)'$	so $\mathbf{D}_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$,
M_l :	$d_t = t$,	$D_{1,t} = 1$	so $\mathbf{D}_1 = 1$,
M_{lc}:	$d_t = t$, but omitted in regression,	$D_{1,t} = 1$	so $\mathbf{D}_1 = 1$,
M_c :	$d_t = 1$,	$D_{1,t} = \emptyset$,	
M_z :	$d_t = \emptyset$,	$D_{1,t} = \emptyset$.	

The cointegration analysis of Johansen (1988, 1995) evolves around the reduced rank restriction

$$\mathsf{H}(r): \quad \operatorname{rank}(\Pi, \Pi_d) \leq r$$

for some $r \leq p$. The reduced rank restriction can be parametised as

$$(\Pi, \Pi_d) = \alpha\beta^{*\prime} \qquad \text{where} \qquad \beta^{*\prime} = (\beta', \delta'),$$

so $\alpha, \beta \in \mathbf{R}^{p\times r}$, $\delta \in \mathbf{R}^{1\times r}$ vary freely. The likelihood ratio test statistic for $\mathsf{H}(r)$ is reviewed below. The interpretation of the hypothesis will, however, depend on the stochastic properties of the process and hence on the parameters. For the standard $\mathsf{I}(1)$-case interpretation is given through the Granger-Johansen representation, see Johansen (1995a, Theorem 4.2). A generalisation of that result is given in §3.

The likelihood ratio test statistic for the rank hypothesis is based on reduced rank regression. Define $X^*_{t-1} = (X'_{t-1}, d_{t-1})'$ or simply as $X^*_{t-1} = X_{t-1}$ if d_{t-1} is omitted from the model equation (2.1). The likelihood is then maximised in two steps. First, ΔX_t and X^*_{t-1} are regressed on the remaining terms giving the least squares residuals

$$(R_{0,t}, R_{1,t}) = \left(\Delta X_t, X^*_{t-1} \middle| \Delta X_{t-1}, \ldots, \Delta X_{t-k}, D_t\right), \tag{2.4}$$

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 sample product moments $S_{ij} = T^{-1}\sum_{t=1}^{T} R_{i,t}R'_{j,t}$ and then solving the eigenvalue problem

$$0 = \det(\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}). \tag{2.5}$$

The likelihood ratio test statistic for the reduced rank restriction $\mathsf{H}(r)$ is given by

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

3 Granger-Johansen representation

To establish a Granger-Johansen representation the rank of the autoregressive level impact matrix Π needs to be known.

Assumption A $\operatorname{rank}(\Pi) = r$.

For the classical $\mathsf{I}(1)$ case the number of unit roots is given as follows.

Assumption B *The number of unit roots is $p - r$.*

The $\mathsf{I}(1)$-condition of Johansen (1988, 1995) is an algebraic condition on the parameters ensuring that the number of unit roots is $p - r$. The next theorem shows that the $\mathsf{I}(1)$-condition holds regardless of the location of the remaining roots.

Theorem 3.1 *Assume A holds. Then Assumption B is equivalent to the condition* $\det(\alpha_\perp' \Psi \beta_\perp) \neq 0$.

Theorem 3.1 implies that under Assumption A, B then $(\beta, \Psi'\alpha_\perp)$ is invertible and so it can serve as a basis for $\mathbb{R}^p$. This is seen by pre-multiplying the basis with $(\overline{\beta}, \beta_\perp)'$ which gives a triangular block matix, see also Johansen (1995a, Exercise 3.7).

The general Granger-Johansen representation theorem now follows. There are two differences in the formulation as compared to Johansen (1995, Theorem 4.2). First, the considered parameter space is larger. Secondly, the representation expresses X_{t-1} in terms of the regressor $Y_{t-1} = (X_{t-1}^{*\prime}\beta^*, \Delta X_{t-1}', \ldots, \Delta X_{t-k+1}')'$ of the model equation (2.1). The result is therefore suited to analysis of the residuals $R_{0,t}$ and $R_{1,t}$. In those respects the result generalises the univariate result of Nielsen (2001, Lemma A1).

Theorem 3.2 *Suppose seasonal deterministic components are excluded,* $D_t = D_{1,t}$. *Assume A, B. Define the process* $Y_t = (X_t^{*\prime}\beta^*, \Delta X_t', \ldots, \Delta X_{t-k+2}')'$, *parameters*

$$C = \beta_\perp(\alpha_\perp'\Psi\beta_\perp)^{-1}\alpha_\perp', \qquad J = \{(I_p - C\Psi)\overline{\beta}, -C\sum_{j=1}^{k-1}\Gamma_j, \ldots, -C\sum_{j=k-1}^{k-1}\Gamma_j\},$$

and the initial condition $\varepsilon_0 = \Psi X_0 + \sum_{\ell=0}^{k-2}\sum_{j=\ell+1}^{k-1}\Gamma_j \Delta X_{-\ell} + \tau_\varepsilon$. *Then*

$$X_t = C\sum_{s=0}^{t}\varepsilon_s + JY_t + \tau_D D_t + \tau_d d_t \qquad \text{with} \qquad Y_t = \mathbf{Y}Y_{t-1} + (\beta, I_p, 0)'\varepsilon_t,$$

where $\mathbf{Y}$ *satisfies* $\det(\mathbf{Y} - I_{\dim \mathbf{Y}}) \neq 0$ *so* $\mathbf{Y}$ *has no roots of unity. In particular, the cointegrating vectors* β *remove unit roots so the relation* $\beta' X_t$ *has no unit roots. The deterministic terms satisfy*

$$\begin{aligned} \tau_d &= C\mu(1,0)' - (I_p - C\Psi)\overline{\beta}\delta', \\ \beta'\tau_D(1,0)' &= \overline{\alpha}'(\Psi C - I)\mu(1,0)' + \overline{\alpha}'(\Psi C\Psi - \Psi)\overline{\beta}\delta', \\ \tau_\varepsilon &= -\mu\{(1,0)'d_{-1} + (I_{(\dim \mathbf{D})-1}, 0)'D_{-1}\}. \end{aligned}$$

In particular it holds $\beta^{*\prime}X_t^* = \beta' X_t + \delta' d_t$ *has no* d_t *component.*

The standard I(1) result of Johansen (1995a, Theorem 4.2) is a special case. This involves the assumption that the $pk - p + r$ roots not of unity are stationary.

Assumption C *The characteric polynomial has* $pk - p + r$ *stationary roots.*

Corollary 3.3 (Johansen, 1995, Theorem 4.2) *Assume A, B, C. Then the process* Y_t *can be given a stationary initial distribution. In particular, the cointegrating relation* $\beta' X_t$ *can be given a stationary initial distribution.*

Other special cases arise under various other assumptions to the $pk-p+r$ roots not of unity. Johansen and Schaumburg (1998) consider the case of seasonal integration. Nielsen (2005) considers the case of co-explosive processes.

When there are additional unit roots the exact representation will depend on the multiplicity of these roots. A result for the standard $\mathsf{I}(2)$ case is given by Johansen (1992). For processes integrated of higher order la Cour (1998) provides a result. Such results are bound to be somewhat involved in terms of notation. A general result tailored towards facilitating the present results is given in §A.5.

4 Asymptotic results

For the asymptotic analysis the normality assumption for the innovations can be replaced by a martingale difference assumption. The assumption is inspired by the analysis of Lai and Wei (1982, 1983). It involves a bound to the conditional moments of the innovations which is used to establish their Marcinkiewic-Zygmund result used for the analysis of the explosive component.

Assumption D *Let $\mathcal{F}_t$ be some filtration so the initial observations are measurable with respect to $\mathcal{F}_0$. Let $(\varepsilon_t, \mathcal{F}_t)$ be a martingale difference assumption, so ε_t is $\mathcal{F}_t$-measurable and $\mathsf{E}(\varepsilon_t|\mathcal{F}_{t-1}) = 0$ a.s. Suppose*
(i) $\sup_t \mathsf{E}(||\varepsilon_t||^{2+\gamma}|\mathcal{F}_{t-1}) < \infty$ *for some* $\gamma > 0$.
(ii) $\mathsf{E}(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}) = \Omega$ *a.s.*

The requirement of constant conditional variance is used for two reasons. First, it is used to establish a Law of Large Number and Functional Central Limit Theorems involving the innovations ε_t and so it can it that respect be replaced by an assumption that such Theorems hold. Secondly, it is used to handle the singular explosive process. Thus, Assumption D could be modified somewhat if singular explosive processes were ruled out.

The first result concerns the consistency of the canonical correlations. This result only requires the cointegration rank is known and does not involve any assumptions to the characteristic roots.

Theorem 4.1 *Assume A, D with $\gamma > 1$. Then*

$$(\hat{\lambda}_1, \ldots, \hat{\lambda}_r) = \mathsf{O}_\mathsf{P}(1), \qquad (\hat{\lambda}_{r+1}, \ldots, \hat{\lambda}_p) = \mathsf{O}_\mathsf{P}(T^{-1}).$$

The proof of Theorem 4.1 includes the notion of a stochastic integral with respect to a singular explosive process. The necessary theory is established in §5.

Strong consistency results can be established for regular vector autoregressions.

Assumption E *The process is regular: any explosive root has geometric multiplicity of one.*

Theorem 4.2 *Assume A, D, E hold with* $\gamma > 0$*. Then*
(*i*) $\liminf_{T\to\infty} \hat{\lambda}_r > 0$ *and* $(\hat{\lambda}_{r+1}, \ldots, \hat{\lambda}_p) = \mathrm{O}(T^{-\xi})$ *a.s. for all* $\xi < \gamma/(2+\gamma)$.
(*ii*) $\liminf_{T\to\infty} \hat{\lambda}_r > 0$ *and* $(\hat{\lambda}_{r+1}, \ldots, \hat{\lambda}_p) = \mathrm{O}(T^{-1}\log T)$ *a.s. if C holds.*

Remark 4.3 *Some strong consistency results can be established when* Y_t *has singular roots. For details see Remarks A.6, A.8 involving an argument of Bauer* (*2009*).

The next result shows that the rank test statistic has the usual asymptotic distribution when the number of unit roots is $p - r$ and singular explosiveness is excluded.

Theorem 4.4 *Consider either of the models* $\mathsf{M}_{lq}, \mathsf{M}_l, \mathsf{M}_{cl}, \mathsf{M}_c, \mathsf{M}_z$*, possible including a seasonal component* $D_{\backslash 1,t}$*. Assume A, B, D, E with* $\gamma > 2$*. Then* LR *has the usual asymptotic distribution described by Johansen* (*1995a, Theorems 6.1, 6.2*):

$$\mathsf{LR} \overset{\mathsf{D}}{\to} \mathrm{tr}\{\textstyle\int_0^1 dB_u F_u' (\int_0^1 F_u F_u' du)^{-1} \int_0^1 F_u dB_u'\}.$$

Here B_u *is a* $(p-r)$*-dimensional standard Brownian motion while* F_u *is given by:*

$$\begin{aligned}
\mathsf{M}_l &: \quad F_u = \{\begin{pmatrix} B_u \\ u \end{pmatrix} \mid 1\}, \quad \mathsf{M}_c : F_u = \begin{pmatrix} B_u \\ 1 \end{pmatrix}, \quad \mathsf{M}_z : F_u = B_u, \\
\mathsf{M}_{lq} &: \quad F_u = \{\begin{pmatrix} (I_{p-r-1}, 0)B_u \\ u \end{pmatrix} \mid 1\} \quad \textit{assuming } \alpha_\perp' \mu_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \neq 0, \\
\mathsf{M}_{cl} &: \quad F_u = \begin{pmatrix} (I_{p-r-1}, 0)B_u \\ 1 \end{pmatrix} \quad \textit{assuming } \alpha_\perp' \mu_1 \neq 0.
\end{aligned}$$

Remark 4.5 *If the process has no explosive components it suffices that* $\gamma > 0$ *in Assumption D as discussed in Remark A.12.*

Special cases of this result are as follows. The standard $\mathsf{I}(1)$ result of Johansen (1995a, Theorem 6.1). The seasonal integration result of Johansen and Schaumburg (1998). The univariate result, $p = 1$, allowing explosive roots by Nielsen (2001). The co-explosive result with one explosive root by Nielsen (2008).

Remark 4.6 *If the Assumptions A, B to the number of unit roots are not satisfied the rank test statistic will not have the correct limit. If the algebraic multiplicity of the unit root is higher than the geometric multiplicity then the process is integrated of order two,* $\mathsf{I}(2)$*, or higher. Johansen* (*1995b*) *and Rahbek, Kongsted and Jørgensen* (*1999*) *discuss the limit distribution in* $\mathsf{I}(2)$ *situations. If the algebraic and geometric* *multiplicity are the same but with more than r unit roots the process is* $\mathsf{I}(1)$*, but with less than r cointegrating relations. The limit distribution is discussed by Nielsen (2004) for a bivariate situation. In general, combinations of such types of distributions can appear.*

Remark 4.7 *For singular vector autoregressions the rank test statistic has the usual limit. As an example consider the bivariate second order vector autoregression*

$$\Delta X_t = \rho \Delta X_{t-1} + \varepsilon_t \qquad \textit{with} \qquad X_0 = \Delta X_0 = 0, \quad \Omega = I_2.$$

Then it holds, see Appendix A.8 for details, that

$$\mathsf{LR} = \mathrm{tr}\{(I - cP)\mathcal{I}_{01}\mathcal{I}_{11}^{-1}\mathcal{I}_{10}\} + \mathrm{o}_{\mathsf{P}}(1),$$

where $\mathcal{I}_{10} = \sum_{t=1}^{T}(\sum_{s=1}^{t-1}\varepsilon_s)\varepsilon_t'$, $\mathcal{I}_{11} = \sum_{t=1}^{T}(\sum_{s=1}^{t-1}\varepsilon_s)^{\otimes 2}$, $c = \rho^2/(\rho^2 - 1)$ *and* $P = w_\perp(w_\perp' w_\perp)^{-1}w_\perp'$ *where* $w_\perp$ *be the orthogonal complement of the Marcinkiewicz-Zygmund limit* $\rho^{-t}\Delta X_t \to w = \sum_{s=1}^{\infty}\rho^{-s}\varepsilon_s$ *a.s.*

5 Convergence of stochastic integrals involving singular explosive processes

The asymptotic analysis of the likelihood statistics involves cross sample moments of random walk type variables and the singular explosive process. To analyse these it is natural to develop some convergence results for stochastic integrals with respect to singular explosive processes. The difficulty is that the singular explosive process is not adapted so the standard semi-martingale result of Jakubowski, Mémin and Pages (1989) does not apply. de Jong and Davidson (2000) considered related stochastic integrals where both the integrand and the integrator are mixing processes with Brownian limits. Here the integrator has to be of a more general type but at the same type it can be exploited that the singular process integrand is a particular nice mixingale.

Consider a singular explosive process $Z_t = \sum_{j=1}^{\infty}\mathbf{W}^{-j}e_{W,t+j}$ where $|\mathrm{eigen}(\mathbf{W})| > 1$ and $e_{W,t}$ is a linear function of ε_t. The simplest stochastic integral of interest is

$$\frac{1}{T}\sum_{t=1}^{T}(\sum_{s=1}^{t-1}\varepsilon_s)Z_{t-1}'.$$

This arises as a cross product sample moment of some of the regressors, hence the timing. Since $(\mathbf{W} - I)Z_{t-1} = e_{W,t} + \Delta Z_t$, see Nielsen (2008, Theorem 3.4), it holds

$$\frac{1}{T}\sum_{t=1}^{T}(\sum_{s=1}^{t-1}\varepsilon_s)Z_{t-1}'(\mathbf{W}' - I) = \frac{1}{T}\sum_{t=1}^{T}(\sum_{s=1}^{t-1}\varepsilon_s)e_{W,t}' + \frac{1}{T}\sum_{t=1}^{T}(\sum_{s=1}^{t-1}\varepsilon_s)(\Delta Z_t)'. \tag{5.1}$$

The first term converges to a stochastic integral in the usual way. The second term is o(1) $a.s.$ To prove this apply partial summation

$$\frac{1}{T}\sum_{t=1}^{T}(\sum_{s=1}^{t-1}\varepsilon_s)(\Delta Z_t)' = \frac{1}{T}(\sum_{t=1}^{T}\varepsilon_t)Z_T' - \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t Z_t'. \tag{5.2}$$

For the first term, $\sum_{t=1}^{T}\varepsilon_t$ is $\mathrm{o}(T^{1/2+\eta})$ for any $\eta > 0$ by a Law of Iterated Logarithms, see Lai and Wei (1985, Theorem 1), while $Z_T = \mathrm{o}(T^{1/2-\eta})$ for a sufficiently small $\eta > 0$, see Nielsen (2008, Corollary 4.3), both assuming D(i) with $\gamma > 0$. The second term can be argued to vanish

These arguments can be generalised for rather general integrands. This is needed because vector autoregressions can generate integrated processes of large order. For this general result introduce the processes

$$J_{T,u} = T^{-1/2}\sum_{t=1}^{\mathrm{int}(Tu)}\varepsilon_t, \quad K_{T,u} = T^{-1/2}\sum_{t=1}^{\mathrm{int}(Tu)} Z_{t-1}$$

defined as $(p + \dim \mathbf{W})$-dimensional process on the space $\mathsf{D}_{\mathbb{R}^{p+\dim \mathbf{W}}}[0, 1]$ of functions on $[0, 1]$ with left limits and right continuity taking the value 0 at 0 endowed with the Skorokhod metric with a univariate deformation. As in (5.1) it holds

$$(\mathbf{W} - I)K_{T,u} = T^{-1/2}\sum_{t=1}^{\mathrm{int}(Tu)} e_{W,t} + T^{-1/2}(Z_{\mathrm{int}(Tu)} - Z_0).$$

The latter term is o(1) uniformly in u with probability one, see Nielsen (2008, Corollary 4.3) assuming D(i). It then holds that

$$(J_T, K_T) \xrightarrow{\mathsf{D}} (J, K), \tag{5.3}$$

on $\mathsf{D}_{\mathbb{R}^{p+\dim \mathbf{W}}}[0, 1]$ by Chan and Wei (1988, Theorem 2.2) assuming D, where the limit is a Brownian motion with variance matrix

$$\mathsf{Var}\begin{pmatrix} J_u \\ K_u \end{pmatrix} = u\begin{pmatrix} \Omega & \Omega_{\varepsilon Z} \\ \Omega_{Z\varepsilon} & \Omega_{ZZ} \end{pmatrix}.$$

Now, let $h : \mathsf{D}_{\mathbb{R}^p}[0, 1] \times [0, 1] \mapsto \mathsf{D}_{\mathbb{R}^m}[0, 1]$ be a continuous function, let τ be the identity function: $\tau(u) = u$, and define $H_T = h(J_T, \tau)$ and $H = h(J, \tau)$. Examples could be integrals like $H_{T,u} = \int_0^u J_{T,s}ds$ representing integrated processes of higher order which will be considered here and powers like $H_{T,u} = J_{T,u}^2$ as well as polynomials in time like $H_{T,u} = u$. Then by the continuous mapping theorem it holds on $\mathsf{D}_{\mathbb{R}^{m+p+\dim \mathbf{W}}}[0, 1]$ that

$$(H_T, J_T, K_T) \xrightarrow{\mathsf{D}} (H, J, K). \tag{5.4}$$

Since J_T is a quadratic martingale with respect to the filtration $\mathcal{F}_{T,u} = \mathcal{F}_{\text{int}(Tu)}$ it holds jointly with (5.4) that $\int_0^u H_{T,s-} dJ'_{T,s} \overset{\mathsf{D}}{\to} \int_0^u H_s dJ'_s$, see Jakubowski, Mémin and Pages (1989). The question is then if $\int_0^u H_{T,s-} dK'_{T,s} \overset{\mathsf{D}}{\to} \int_0^u H_s dK'_s$, jointly with the previous convergence. The difficulty is that K_T is not $\mathcal{F}_T$-adapted. The solution is to decompose

$$\begin{aligned} & \int_0^u H_{T,s-} dK'_{T,s}(\mathbf{W} - I) \\ = \quad & \int_0^u H_{T,s-} d(T^{-1/2} \sum_{t=1}^{\text{int}(Ts)} e_{W,t})' + \int_0^u H_{T,s-} dT^{-1/2}(Z_{\text{int}(Ts)} - Z_0)'. \end{aligned} \tag{5.5}$$

Assuming that H_T is $\mathcal{F}_T$-adapted the first term converges by Jakubowski, Mémin and Pages (1989) so it is left to argue that the latter vanishes.

Theorem 5.1 *Assuming D with $\gamma > 0$, that h satisfy the condition*

$$\sup_{0 \le u \le 1} || \frac{\partial}{\partial u} \{h(\cdot, \cdot)(u)\} \Big|_{(J_T, \tau)} || \overset{a.s.}{=} \mathrm{o}(T^{\eta})$$

for all $\eta > 0$ and H_T is $\mathcal{F}_T$-adapted then it holds on $\mathsf{D}_{\mathbb{R}^{(m+1)(p+\dim \mathbf{W}+1)-1}}[0,1]$ that

$$(H_T, J_T, K_T, \int_0^{\cdot} H_{T,s-} dJ'_{T,s}, \int_0^{\cdot} H_{T,s-} dK'_{T,s}) \overset{\mathsf{D}}{\to} (H, J, K, \int_0^{\cdot} H_s dJ'_s, \int_0^{\cdot} H_s dK'_s).$$

Remark 5.2 *The condition to the derivative of h holds if $h(J_T, \tau)$ satisfies a Law of Iterated Logarithms. This holds for (repeated) integrals of random walks, see Lai and Wei (1985, Theorem 1) of Nielsen (2005, Theorem 5.1) and for power functions due to the same Law of Iterated Logarithms.*

A Proofs

A.1 Proofs of representation results

Proof of Theorem 3.1. For this argument the deterministic terms are irrelevant. Let $Y_t = \{(\beta' X_t)', \Delta X_t', \ldots, \Delta X_{t-k+2}'\}'$. The companion vector $(X_t'\beta_\perp, Y_t')'$ satisfies

$$\begin{pmatrix} \beta_\perp' X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} I_{p-r} & \beta_\perp'\nu \\ 0 & \mathbf{Y} \end{pmatrix} \begin{pmatrix} \beta_\perp' X_{t-1} \\ Y_{t-1} \end{pmatrix} + \begin{pmatrix} \beta_\perp' \\ \iota_Y \end{pmatrix} \varepsilon_t,$$

where $\mathbf{Y}, \iota_Y, \nu$ are given below. The triangular structure of the companion matrix implies that Assumption B is equivalent to $\det(\mathbf{Y} - I_{p(k-1)+r}) \neq 0$. Now, $\mathbf{Y}, \iota_Y, \nu$ are

$$\mathbf{Y} = \begin{pmatrix} I_r + \beta'\alpha & \beta'\Gamma_1 & \beta'\varphi \\ \alpha & \Gamma_1 & \varphi \\ 0 & \psi & N_0 \end{pmatrix}, \quad \iota_Y = \begin{pmatrix} \beta' \\ I_p \\ 0 \end{pmatrix}, \quad \nu = (\alpha, \Gamma_1, \ldots, \Gamma_{k-1}) \qquad \text{(A.1)}$$

where $\varphi = (\Gamma_2, \ldots, \Gamma_{k-1})$ and $\psi' = (I_p, 0)$ are $\{p \times p(k-2)\}$-dimensional, while the $\{p(k-2) \times p(k-2)\}$-dimensional N_x and its inverse, for $x \neq 0$, are

$$N_x = \begin{pmatrix} -xI_p & & & \\ I_p & \ddots & & \\ & \ddots & \ddots & \\ & & I_p & -xI_p \end{pmatrix}, \quad N_x^{-1} = -\begin{pmatrix} x^{-1}I_p & & \\ \vdots & \ddots & \\ x^{1-k}I_p & \cdots & x^{-1}I_p \end{pmatrix}^{-1}.$$

Partitioned inversion gives $\det(\mathbf{Y} - I_{kp-p+r}) = \det(N_1)\det(\mathcal{D})$, where $\det(-N_1) = 1$ and $\mathcal{D}$ is given by

$$\mathcal{D} = \begin{pmatrix} \beta'\alpha & \beta'\Gamma_1 \\ \alpha & \Gamma_1 - I_p \end{pmatrix} - \begin{pmatrix} \beta' \\ I_p \end{pmatrix} \varphi N_1^{-1} \psi \begin{pmatrix} 0 & I_p \end{pmatrix}.$$

Inserting the expressions for φ, N_1^{-1}, ψ and recalling $\Psi = I - \sum_{j=1}^{k-1} \Gamma_j$ gives

$$\mathcal{D} = \begin{pmatrix} \beta'\alpha & \beta'\Gamma_1 \\ \alpha & \Gamma_1 - I_p \end{pmatrix} + \begin{pmatrix} \beta' \\ I_p \end{pmatrix} (0, \sum_{j=2}^{k-1} \Gamma_j) = \begin{pmatrix} \beta'\alpha & \beta'(I_p - \Psi) \\ \alpha & -\Psi \end{pmatrix}.$$

Pre-multiply and post-multiply $\mathcal{D}$ with regular matrices

$$\left\{ \begin{matrix} 0 & \begin{pmatrix} \overline{\alpha}' \\ \alpha_\perp' \end{pmatrix} \\ I_r & -\beta' \end{matrix} \right\} \mathcal{D} \left\{ \begin{matrix} 0 & I_r \\ (\overline{\beta}, \beta_\perp) & 0 \end{matrix} \right\} = \begin{pmatrix} -\overline{\alpha}'\Psi\overline{\beta} & -\overline{\alpha}'\Psi\beta_\perp & I_r \\ -\alpha_\perp'\Psi\overline{\beta} & -\alpha_\perp'\Psi\beta_\perp & 0 \\ I_r & 0 & 0 \end{pmatrix}.$$

The latter matrix is regular if and only if $\det(\alpha_\perp'\Psi\beta_\perp) \neq 0$ as desired. ■

Proof of Theorem 3.2. *Homogenous equation.* Leaving out deterministic terms and recalling that $\Psi = I_p - \sum_{j=1}^{k-1} \Gamma_j$ the model equation (2.1) can be rewritten as

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

Insert $\Delta X_{t-j} - \Delta X_t = -\sum_{\ell=0}^{j-1} \Delta^2 X_{t-\ell}$ and interchange the two sums to get

$$\Psi \Delta X_t = \alpha\beta' X_{t-1} - \sum_{\ell=0}^{k-2} (\sum_{j=\ell+1}^{k-1} \Gamma_j) \Delta^2 X_{t-\ell} + \varepsilon_t.$$

Pre-multiply by $C = \beta_\perp (\alpha'_\perp \Psi \beta_\perp)^{-1} \alpha'_\perp$, sum over t, and recall the definition of ε_0 to get

$$C\Psi X_t = \sum_{s=0}^{t} C\varepsilon_s - \sum_{\ell=0}^{k-2} (\sum_{j=\ell+1}^{k-1} C\Gamma_j) \Delta X_{t-\ell}. \tag{A.2}$$

Assuming A, B then Theorem 3.1 implies that $(\beta, \Psi'\alpha_\perp)$ is a basis. Then $C\Psi$ is the associated skew projection on $\beta_\perp$ along $\alpha'_\perp \Psi$, and it holds $(I_p - C\Psi)\beta_\perp = 0$ giving the skew projection identity

$$I_p = C\Psi + (I_p - C\Psi)\overline{\beta}\beta'. \tag{A.3}$$

Therefore $X_t = C\Psi X_t + (I_p - C\Psi)\overline{\beta}\beta' X_t$. Insert $C\Psi X_t$ from (A.2) and identify the JY_t component from the $\Delta X_{t-\ell}$ and $\beta' X_t$ terms. Note that Y_t has no unit roots under Assumption B as discussed in the proof of Theorem 3.1.

Inhomogenous equation. It is assumed that the seasonal deterministic terms $D_{\setminus 1,t}$ are absent so $D_t = D_{1,t}$. Replace ε_t by $\varepsilon_t + \alpha\delta' d_{t-1} + \mu D_{t-1}$. For the common trend component, $C \sum_{s=0}^{t} \varepsilon_s$, the additional contribution is $C\mu \sum_{s=0}^{t} D_{s-1}$. Since d_t and D_t satisfy the equation (2.2) then

$$(1, 0) D_{t-1} = \Delta d_t \quad \text{and} \quad (0, I_{(\dim \mathbf{D})-1}) D_{t-1} = (I_{(\dim \mathbf{D})-1}, 0) \Delta D_t. \tag{A.4}$$

The determistic contribution to the common trends is then

$$C\mu \sum_{s=0}^{t} D_{s-1} = C\mu\, (1, 0)'\, (d_t - d_{-1}) + C\mu (I_{(\dim \mathbf{D})-1}, 0)' (D_t - D_{-1}). \tag{A.5}$$

The equation for the non-unit root component is, in terms of $\mathbf{Y}, \iota_Y$ of (A.1),

$$Y_t = \mathbf{Y} Y_{t-1} + \iota_Y (\varepsilon_t + \mu D_{t-1}),$$

noting that Y_t now includes the component $\beta^{*\prime} X_t^* = \beta' X_t + \delta' d_t$. Since $\mathbf{Y}$ does not have unit roots then $\tilde{Y}_t = Y_t - \kappa_D D_t$ solves the homogeneous equation for Y_t for some κ_D as argued in Nielsen (2005, §3).

Combining these results it follows that

$$X_t = \tilde{X}_t + \tau_D D_t + \tau_d d_t + \tilde{\tau}_\varepsilon, \tag{A.6}$$

where $\tilde{X}_t$ solves the homogenous equation. It holds $\tau_d = C\mu(1,0)' - (I_p - C\Psi)\overline{\beta}\delta'$ where the first term originates from the common trend, while the second term arises from $\beta^{*\prime} X_t^*$ with impact $(I_p - C\Psi)\overline{\beta}$ noting that ΔX_t has no d_t term. Further, $\tilde{\tau}_\varepsilon = C\tau_\varepsilon$ where $\tau_\varepsilon = -\mu\{(1,0)' d_{-1} + (I_{(\dim \mathbf{D})-1}, 0)' D_{-1}\}$ is the initial condition for the common trends, see (A.5), noting that initial values for Y_t are implicitly included in the equation for Y_t. The term τ_D is to be determined.

Insert the expression (A.6) for X_t in the model equation (2.1) to get

$$\begin{aligned}\{\Delta\tilde{X}_t + \tau_D \Delta D_t + \tau_d (1,0) D_{t-1}\} &= \alpha(\beta'\tilde{X}_{t-1} + \beta'\tau_D D_{t-1}) \\ &+ \sum_{j=1}^{k-1} \Gamma_j \{\Delta\tilde{X}_{t-j} + \tau_D \Delta D_{t-j} + \tau_d(1,0) D_{t-j-1}\} + \mu D_{t-1} + \varepsilon_t,\end{aligned}$$

noting $\beta'\tau_d = -\delta'$. As $\tilde{X}_t$ solves the homogeneous equation it must hold that

$$\tau_D \Delta D_t + \tau_d(1,0) D_{t-1} = \alpha\beta'\tau_D D_{t-1} + \sum_{j=1}^{k-1} \Gamma_j \{\tau_D \Delta D_{t-j} + \tau_d (1,0) D_{t-j-1}\} + \mu D_{t-1}.$$

Pre-multiply by β', focus on the first element of D_t, that is $D_t^{(1)} = (1,0) D_t$ say, and note that the first element of ΔD_t does not involve $D_t^{(1)}$. Hence, $\beta'\tau_D(1,0)'$ solves

$$\tau_d = \alpha\beta'\tau_D(1,0)' + \sum_{j=1}^{k-1} \Gamma_j \tau_d + \mu(1,0)'.$$

Insert the expression for τ_d, rearrange and pre-multiply with $\overline{\alpha}'$ to get the desired expression for $\beta'\tau_D(1,0)'$. ■

A.2 Some initial remarks on the eigenvalue problem

The cointegration analysis is done in terms of the residuals $R_{0,t}$ and $R_{1,t}$ defined in (2.4). These residuals arise by regressing on $\Delta X_{t-1}, \ldots, \Delta X_{t-k+1}, D_{t-1}$. As indicated by the model equation (2.1) and the Granger-Johansen representation in Theorem 3.2 then it is convenient to extend this set of regressors by $\beta^{*\prime} X_{t-1}^*$ giving the regressor $(Y_{t-1}', D_{t-1}')'$ where $Y_t = (X_t^{*\prime}\beta^*, \Delta X_t', \ldots, \Delta X_{t-k+2}')'$.

To appreciate the consequences of this extention the residual $R_{1,t}$ has to rotated by β^* as well as a complement, $\beta_\perp^*$ say, of β^*, so $(\beta^*, \beta_\perp^*)$ has full rank, but it is not necessary that $\beta_\perp^{*\prime}\beta^* = 0$. Different choices for $\beta_\perp^*$ depending on whether B is assumed or not, see §A.4, A.5. Thus, define the residuals

$$R_{\beta,t} = \beta^{*\prime} R_{1,t}, \quad R_{0\cdot\beta,t} = (\Delta X_t \mid Y_{t-1}, D_{t-1}), \quad R_{\beta_\perp\cdot\beta,t} = (\beta_\perp^{*\prime} X_{t-1}^* \mid Y_{t-1}, D_{t-1}),$$

noting that by the model equation (2.1) then $R_{0\cdot\beta,t} = (\varepsilon_t \mid Y_{t-1}, D_{t-1})$. Define also the conditional product moments

$$\begin{pmatrix} S_{00\cdot\beta} & S_{0\beta_\perp\cdot\beta} \\ S_{\beta_\perp 0\cdot\beta} & S_{\beta_\perp\beta_\perp\cdot\beta} \end{pmatrix} = T^{-1}\sum_{t=1}^{T} \begin{pmatrix} R_{0\cdot\beta,t} \\ R_{\beta_\perp\cdot\beta,t} \end{pmatrix}^{\otimes 2}.$$

The original eigenvalue problem $0 = \det(\lambda S_{00} - S_{01}S_{11}^{-1}S_{10})$ can then be written as

$$0 = \det\{\lambda S_{00\cdot\beta} + (\lambda-1)S_{0\beta}S_{\beta\beta}^{-1}S_{\beta 0} - S_{0\beta_\perp\cdot\beta}S_{\beta_\perp\beta_\perp\cdot\beta}^{-1}S_{\beta_\perp 0\cdot\beta}\}. \tag{A.7}$$

The asymptotic analysis of the cointegration rank test then rests on an analysis of the terms $S_{00\cdot\beta}$, $S_{0\beta}S_{\beta\beta}^{-1}S_{\beta 0}$, $S_{0\beta_\perp\cdot\beta}S_{\beta_\perp\beta_\perp\cdot\beta}^{-1}S_{\beta_\perp 0\cdot\beta}$. The first two terms involve only the extended regressor $(Y_t', D_t')'$ which is a generalised cointegration vector. These terms are discussed in §A.3. Two different analyses are made for the third term depending on whether B is assumed or not, see §A.4, §A.5.

For the analysis the following algebraic result will be useful.

Lemma A.1 *Define ε_t and $x_t = (y_t', z_t')'$. Then $S_{\varepsilon y\cdot z}S_{yy\cdot z}^{-1}S_{y\varepsilon\cdot z}$ and $S_{\varepsilon z}S_{zz}^{-1}S_{z\varepsilon}$ are both* $\mathrm{O}(S_{\varepsilon x}S_{xx}^{-1}S_{x\varepsilon})$.

Proof of Lemma A.1. By the partial inversion formula then

$$S_{\varepsilon y\cdot z}S_{yy\cdot z}^{-1}S_{y\varepsilon\cdot z} + S_{\varepsilon z}S_{zz}^{-1}S_{z\varepsilon} = S_{\varepsilon x}S_{xx}^{-1}S_{x\varepsilon}.$$

Then apply that all involved terms are positive semi-definite. ■

A.3 Analysis of the generalised cointegration vector

The terms $S_{00\cdot\beta}$, $S_{0\beta}S_{\beta\beta}^{-1}S_{\beta 0}$ are investigated.

The extended regressor $(Y_t', D_t')'$ where $Y_t = (X_t^{*\prime}\beta^*, \Delta X_t', \ldots, \Delta X_{t-k+2}')'$ satisfies

$$\begin{pmatrix} Y_t \\ D_t \end{pmatrix} = \begin{pmatrix} \mathbf{Y} & \iota_Y\mu \\ 0 & \mathbf{D} \end{pmatrix}\begin{pmatrix} Y_{t-1} \\ D_t \end{pmatrix} + \begin{pmatrix} \iota_Y \\ 0 \end{pmatrix}\varepsilon_t,$$

where $\mathbf{Y}, \iota_Y$ were given in (A.1). Following the argument in Nielsen (2005, §3) an invertible matrix M and a matrix m exist so

$$\begin{pmatrix} M & m \\ 0 & I \end{pmatrix}\begin{pmatrix} Y_t \\ D_t \end{pmatrix} = \begin{pmatrix} K_t \\ W_t \\ D_t \end{pmatrix} \tag{A.8}$$

satisfies the equation

$$\begin{pmatrix} K_t \\ W_t \\ D_t \end{pmatrix} = \begin{pmatrix} \mathbf{K} & 0 & \mu_K \\ 0 & \mathbf{W} & 0 \\ 0 & 0 & \mathbf{D} \end{pmatrix}\begin{pmatrix} K_{t-1} \\ W_{t-1} \\ D_{t-1} \end{pmatrix} + \begin{pmatrix} M\iota_Y \\ 0 \end{pmatrix}\varepsilon_t.$$

where $|\text{eigen}(\mathbf{K})| \leq 1$ and $|\text{eigen}(\mathbf{W})| > 1$. In fact, M can be chosen so

$$\mathbf{K} = \begin{pmatrix} \mathbf{U} & & \\ & \mathbf{V}_1 & \\ & & \mathbf{V}_{\backslash 1} \end{pmatrix}, \qquad K_t = \begin{pmatrix} U_t \\ V_{1,t} \\ V_{\backslash 1,t} \end{pmatrix},$$

where $|\text{eigen}(\mathbf{U})| < 1$, $|\text{eigen}(\mathbf{V}_{\backslash 1})| = 1$ so $\text{eigen}(\mathbf{V}_{\backslash 1}) \neq 1$, and $\text{eigen}(\mathbf{V}_1) = 1$. If $\mathbf{K}$, $\mathbf{D}$ have no common eigenvalues m could be chosen so also $\mu_K = 0$.

Some further analysis is needed for the explosive component. This satisfies

$$W_t = \mathbf{W} W_{t-1} + e_{W,t},$$

where $e_{W,t}$ are the elements of $M \iota_Y \varepsilon_t$ associated with W_t. The Marcinkiewicz-Zygmund result of Lai and Wei (1983) then shows

$$\mathbf{W}^{-t} W_t \overset{a.s.}{\rightarrow} W = W_0 + Z_0 \quad \text{where} \quad Z_t = \sum_{j=1}^{\infty} \mathbf{W}^{-j} e_{W,t+j},$$

where W has a continuous distribution assuming D. As pointed out by Anderson (1959) then $\mathbf{W}^t W$ may have linearly dependent elements which will give a singularity that needs to be taken into account in the asymptotic analysis. Nielsen (2008, Theorem 3.1) shows that this singularity arises when some of the eigenvalues of $\mathbf{W}$ have geometric multiplicity larger than one. The degree of singularity is determined by the dimension n which is the sum of the dimensions of the largest Jordan blocks associated with the distinct eigenvalues of $\mathbf{W}$. Moreover, W_t has the representation

$$W_t = w\lambda_t - Z_t, \tag{A.9}$$

where $w \in \mathbb{R}^{\dim \mathbf{W} \times n}$ is a function of the limiting random vector W and has full column rank with probability one, while the vector $\lambda_t \in \mathbb{R}^n$ is deterministic and of exponential order in t; see (A.16) for an example.

Having the singularity in mind the process Y_t can be decomposed a little further. Define the random transformation

$$N = \begin{pmatrix} N_Q \\ w'_{\perp} N_W \end{pmatrix} \quad \text{where} \quad N_Q = \begin{pmatrix} I_{\dim Y_t - \dim \mathbf{W}} & 0 \\ 0 & \overline{w}' \end{pmatrix}, \quad N_W = \begin{pmatrix} 0 \\ I_{\dim \mathbf{W}} \end{pmatrix}',$$

so N is invertible with probability one. A process Q_t exists so

$$\begin{pmatrix} Q_t \\ w'_{\perp} Z_t \end{pmatrix} = NMY_t, \tag{A.10}$$

and Q_t satisfies $Q_t = \mathbf{Q} Q_{t-1} + N_Q(\mu'_K, 0)' D_{t-1} + N_Q M \iota_Y \varepsilon_t$ where $\mathbf{Q}$ has non-explosive and regularly explosive eigenvalues.

The singular component Z_{t-1} satisfies jointly with ε_t a Law of Large Numbers

$$\frac{1}{T}\sum_{t=1}^{T}\left(\begin{array}{c} Z_{t-1} \\ \varepsilon_t \end{array}\middle| D_{t-1}\right)^{\otimes 2} \overset{a.s.}{\to} \left(\begin{array}{cc} \Omega_{ZZ} & \Omega_{Z\varepsilon} \\ \Omega_{\varepsilon Z} & \Omega \end{array}\right), \tag{A.11}$$

see Nielsen (2008, Equation 4.5, Theorem 4.7) assuming A, D, where

$$\Omega_{ZZ} = \sum_{j=1}^{\infty} \mathbf{W}^{-j} N_W M \iota_Y \Omega \iota_Y' M' N_W' (\mathbf{W}')^{-j}, \quad \Omega_{Z\varepsilon} = \mathbf{W}^{-1} N_W M \iota_Y \Omega.$$

In singular situations some bias terms arise with the following properties.

Lemma A.2 *Define the terms*

$$\begin{aligned} \Omega_{\varepsilon\varepsilon\cdot Z} &= \Omega - \Omega_{\varepsilon Z} w_\perp (w_\perp' \Omega_{ZZ} w_\perp)^{-1} w_\perp' \Omega_{Z\varepsilon}, \\ \alpha_{\text{lim}} &= \alpha + \alpha_{\text{bias}} \quad \textit{where} \quad \alpha_{\text{bias}} = \{0_{p\times \dim Q}, \Omega_{\varepsilon Z} w_\perp (w_\perp' \Omega_{ZZ} w_\perp)^{-1} w_\perp'\} N_W M \iota_Y. \end{aligned}$$

(i) Assuming D then $\Omega_{\varepsilon\varepsilon\cdot Z}$ and $\alpha_{\text{lim}}'\alpha_{\text{lim}}$ are invertible a.s.
(ii) Assuming E then $\Omega_{\varepsilon\varepsilon\cdot Z} = \Omega$ and $\alpha_{\text{lim}} = \alpha$.

Proof of Lemma A.2. This follows from Nielsen (2008, Lemma A.2). ■

Finally, it is convenient to define the residuals and product moment matrices

$$R_{Q,t} = (Q_{t-1} \mid D_{t-1}), \quad R_{(\varepsilon,Z),t} = \{\left(\begin{array}{c} \varepsilon_t \\ Z_{t-1} \end{array}\right) \mid D_{t-1}\}, \quad S_{ij} = T^{-1}\sum_{t=1}^{T} R_{i,t}R_{j,t}',$$

for $i, j = Q, (\varepsilon, Z)$. Some weak and strong convergence results are established for S_{ij}.

Lemma A.3 *Assuming A, D with $\gamma > 1$ then*
(i) $S_{QQ}^{-1} = \mathsf{O}_\mathsf{P}(1)$,
(ii) $S_{QQ}^{-1/2} S_{Q,(\varepsilon,Z)} S_{(\varepsilon,Z),(\varepsilon,Z)}^{-1/2} = \mathsf{o}_\mathsf{P}(1)$.

Proof of Lemma A.3. *(i)* The components K_t, λ_t, $w_\perp' Z_t$ of Q_t are uncorrelated in probability due to Nielsen (2005, Theorem 9.1, 9.2, 9.4), Nielsen (2008, Theorem 4.7). Then apply Nielsen (2008, Theorem 4.9) for each element.

(ii) Follows from Nielsen (2008, Theorem 4.7). Note that if there are no singular component then $\gamma > 0$ suffices in *(ii)* using Nielsen (2005, Theorem 2.4) instead. ■

Lemma A.4 *Assuming A, D with $\gamma > 1$ then*
(i) $S_{00\cdot\beta} = \Omega_{\varepsilon\varepsilon\cdot Z} + \mathsf{o}_\mathsf{P}(1)$.
(ii) $S_{0\beta} S_{\beta\beta}^{-1} = \alpha_{\text{lim}} + \mathsf{o}_\mathsf{P}(1)$.
(iii) $S_{\beta\beta}^{-1} = \mathsf{O}_\mathsf{P}(1)$.

Proof of Lemma A.4. (i) Note $R_{0,t} = (\varepsilon_t | Y_{t-1}, D_{t-1})$. Transform Y_t by NM so $R_{0,t} = (\varepsilon_t | Q_{t-1}, w'_\perp Z_{t-1}, D_{t-1})$. By the uncorrelatedness of Q_{t-1} and $(\varepsilon'_t, Z'_{t-1})'$, see Lemma A.3$(ii)$ assuming A, D then $S_{00\cdot\beta} = T^{-1}\sum_{t=1}^{T}(\varepsilon_t | w'_\perp Z_{t-1})^{\otimes 2} + \mathrm{o_P}(1)$. Then use the Law of Large Numbers in (A.11) assuming A, D.

(ii) Under the hypothesis the model equation implies $S_{0\beta}S_{\beta\beta}^{-1} = \alpha + S_{\varepsilon\beta}S_{\beta\beta}^{-1}$ where the latter term is the partial regression estimator

$$S_{\varepsilon\beta}S_{\beta\beta}^{-1} = S_{\varepsilon Y}S_{YY}^{-1}\iota_\beta \quad \text{with} \quad R_{Y,t} = (Y_{t-1}|D_{t-1}),$$

and where $\iota_\beta = \{I_r, 0_{r\times p(k-1)}\}'$. As in (i) transform Y_t by NM and apply Lemma A.3(ii) to get

$$S_{\varepsilon\beta}S_{\beta\beta}^{-1} = (S_{\varepsilon Q}S_{QQ}^{-1}, S_{\varepsilon Z}S_{ZZ}^{-1})NM\iota_\beta\{1 + \mathrm{o_P}(1)\},$$

defined in terms of $R_{Q,t} = (Q_{t-1}|D_{t-1})$ and $R_{Z,t} = (w'_\perp Z_{t-1}|D_{t-1})$. For the first term note $S_{\varepsilon Q}S_{QQ}^{-1/2} = \mathrm{o}(1)$ *a.s.* by Nielsen (2005, Theorem 2.4), while $S_{QQ}^{-1} = \mathrm{O_P}(1)$ by Lemma A.3(i). For the second term use the Law of Large Numbers in (A.11).

(iii) As in (ii) note $S_{\beta\beta} = \iota'_\beta S_{YY}\iota_\beta$, transform Y_t by NM, use Lemma A.3 to get

$$S_{\beta\beta} = \iota'_\beta(M')^{-1}(N')^{-1}(S_{QQ}, S_{ZZ})N^{-1}M^{-1}\iota_\beta\{1 + \mathrm{o_P}(1)\}.$$

Since M, N have full rank with probability one then by the Poincaré separation theorem, see Magnus and Neudecker (1988, Theorem 11.12) it suffices to argue that $\lambda_{\min}\{T^{-1}\sum_{t=1}^{T}(Q_{t-1}, w'_\perp Z_{t-1}|D_{t-1})^{\otimes 2}\}$ and $\lambda_{\min}\{T^{-1}\sum_{t=1}^{T}(w_\perp Z_{t-1}|D_{t-1})^{\otimes 2}\}$ have positive limiting points. The latter follows from the Law of Large Numbers in (A.11), while $S_{QQ}^{-1} = \mathrm{O_P}(1)$ by Lemma A.3(i). ■

Lemma A.5 *Assuming A, D, E with* $\gamma > 0$ *then*
(i) $S_{00\cdot\beta} \to \Omega_{\varepsilon\varepsilon}$ *a.s.,*
(ii) $S_{0\beta}S_{\beta\beta}^{-1} = \alpha + \mathrm{o}(1)$ *a.s.,*
(iii) $\liminf \lambda_{\min}(S_{\beta\beta}) > 0$ *a.s.*

Remark A.6 *The results in Lemma A.5 hold more generally. An argument could be made along the lines of Lemma A.3, A.4 under either of the following conditions:*
(a) *If the model has singular explosive terms and deterministic terms, but* Y_t *has no roots on the unit circle and Assumption D holds with* $\gamma > 1$*; see Nielsen (2008, Theorem 4.7).*
(b) *If the model has singular explosive terms, but no deterministic terms, and Assumption D holds with* $\gamma > 1$*, then* Y_t *can have roots on the unit circle at rational frequencies* $\exp(\pm i2\pi p/q)$ *where* $p, q \in \mathbb{N}$ *so* $0 < 2p < q$ *as long as these roots have the same algebraic and geometric multiplicity. This is argued by combining Nielsen (2008, Theorem 4.7) with Bauer (2009).*

Proof of Lemma A.5. (i, ii) This is proved in the same way as Nielsen (2005, Theorem 2.4, Corollary 2.6, Theorem 2.8). Those results are concerned with vector autoregressions so an adjustment has to be made since the regressor Y_{t-1} only is a part of the companion vector of a vector autoregression.

(iii) Combine Nielsen (2005, Corollary 9.5) with the argument involving the Poincaré separation theorem in Lemma A.4(iii). ■

A.4 Analysis of the generalised common trends assuming B

When Assumptions A, B hold the Granger-Johansen representation in Theorem 3.2 applies and the analysis of $R_{\beta_\perp \cdot \beta, t}$ is relatively simple. As complement of β^* chose

$$\beta_\perp^{*\prime} = \mathcal{B} \begin{pmatrix} \alpha_\perp' \Psi \beta_\perp \overline{\beta}_\perp' & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} I_p & -\tau_d \\ 0 & 1 \end{pmatrix}, \tag{A.12}$$

so $(\beta^*, \beta_\perp^*)$ is regular, but $\beta_\perp^{*\prime}\beta^*$ need not be zero. Here $\mathcal{B} = I_{p-r+1}$ if d_{t-1} is present in the model equation (2.1) whereas if d_{t-1} is absent then $\alpha_\perp' \Psi \beta_\perp \overline{\beta}_\perp' \tau_d = \alpha_\perp' \mu_1 (1, 0)'$ so $\mathcal{B} = \{I_{p-r}, \alpha_\perp' \mu_1 (1, 0)'\}$. Combing the representation of Theorem 3.2 along with $\beta_\perp^{*\prime}$ then gives, for instance if d_{t-1} is present, that

$$R_{\beta_\perp \cdot \beta, t} = \mathcal{B}\{ \begin{pmatrix} \alpha_\perp' \sum_{s=0}^{t-1} \varepsilon_s \\ d_{t-1} \end{pmatrix} \mid Y_{t-1}, D_{t-1}\}. \tag{A.13}$$

Two results then emerge concerning $S_{0\beta_\perp \cdot \beta} S_{\beta_\perp \beta_\perp \cdot \beta}^{-1} S_{\beta_\perp 0 \cdot \beta}$. The first is a consistency result and the second a distributional result.

Lemma A.7 *Assume A, D with* $\gamma > 0$. *Then*
(i) *If C holds then* $S_{0\beta_\perp \cdot \beta} S_{\beta_\perp \beta_\perp \cdot \beta}^{-1} S_{\beta_\perp 0 \cdot \beta} = \mathrm{O}(T^{-1} \log T)$ *a.s.,*
(ii) *If E holds then* $S_{0\beta_\perp \cdot \beta} S_{\beta_\perp \beta_\perp \cdot \beta}^{-1} S_{\beta_\perp 0 \cdot \beta} = \mathrm{O}(T^{-\xi})$ *a.s. for all* $\xi < \gamma/(2+\gamma)$.

Remark A.8 *The results in Lemma A.7 hold more generally under the conditions* $(a), (b)$ *of Remark A.6. The proof would be a modification of the proof of Lemma A.10. There are two arguments. First, the uncorrelatedness of* Y_{t-1} *and* $\sum_{s=1}^{t-1} \varepsilon_s$ *also hold with explosive roots under conditions* $(a), (b)$ *so* $S_{\beta_\perp \beta_\perp \cdot \beta} = S_{CC}\{1 + \mathrm{o}(1)\}$ *with* $S_{CC} = T^{-1} \sum_{t=1}^{T} (\alpha_\perp' \sum_{s=1}^{t-1} \varepsilon_s)^{\otimes 2}$. *Secondly, the uncorrelatedness of* Q_{t-1} *and* (ε_t, Z_{t-1}) *then shows* $S_{0\beta_\perp \cdot \beta} S_{CC}^{-1/2} = T^{-1} \sum_{t=1}^{T} (\varepsilon_t | w_\perp' Z_{t-1})(c\alpha_\perp' \sum_{s=1}^{t-1} \varepsilon_s)' S_{CC}^{-1/2} + \mathrm{o}(1)$. *Then argue as in Remark A.6.*

Proof of Lemma A.7. Note $R_{0,t} = (\varepsilon_t | Y_{t-1}, D_{t-1})$. Let $S_t = (X_t^{*\prime} \beta_\perp^*, d_t, Y_t', D_t')$ be the full companion vector. Then $S_{0\beta_\perp \cdot \beta} S_{\beta_\perp \beta_\perp \cdot \beta}^{-1} S_{\beta_\perp 0 \cdot \beta} = \mathrm{O}(S_{\varepsilon S} S_{SS}^{-1} S_{S\varepsilon})$ by Lemma A.1. Then apply Nielsen (2005, Theorem 2.4) assuming either of C, E. ■

Lemma A.9 *Consider either of the models* $\mathsf{M}_{lq}, \mathsf{M}_l, \mathsf{M}_{cl}, \mathsf{M}_c, \mathsf{M}_z$. *Assume A, B, D, E with* $\gamma > 0$. *Then*
(i) $S_{0\beta}S_{\beta\beta}^{-1} = \alpha + S_{\varepsilon\beta}S_{\beta\beta}^{-1}$ *where* $S_{\varepsilon\beta}S_{\beta\beta}^{-1}S_{\beta\varepsilon} = \mathrm{o_P}(T^{-\xi})$ *for all* $\xi < \gamma/(2+\gamma)$.
(ii) $S_{0\beta_\perp\cdot\beta}S_{\beta_\perp\beta_\perp\cdot\beta}^{-1}S_{\beta_\perp 0\cdot\beta} = \mathrm{O_P}(T^{-1})$.
(iii) *Define* B_u, F_u *as in Theorem 4.4, let* $\Omega_{\alpha_\perp\alpha_\perp} = \alpha_\perp'\Omega\alpha_\perp$ $\tilde{B}_u$. *Then*

$$T\Omega_{\alpha_\perp\alpha_\perp}^{-1/2}\alpha_\perp' S_{0\beta_\perp\cdot\beta}S_{\beta_\perp\beta_\perp\cdot\beta}^{-1}S_{\beta_\perp 0\cdot\beta}\alpha_\perp\Omega_{\alpha_\perp\alpha_\perp}^{-1/2} \overset{\mathsf{D}}{\to} \mathbf{B} = \textstyle\int_0^1 dB_u F_u'(\int_0^1 F_u F_u' du)^{-1}\int_0^1 F_u dB_u'.$$

Proof of Lemma A.9. (i) Under the hypothesis the model equation implies $S_{0\beta}S_{\beta\beta}^{-1} = \alpha + S_{\varepsilon\beta}S_{\beta\beta}^{-1}$. To establish the desired bound note that by Lemma A.1 then $S_{\varepsilon\beta}S_{\beta\beta}^{-1}S_{\beta\varepsilon} = \mathrm{O}(S_{\varepsilon Y}S_{YY}^{-1}S_{Y\varepsilon})$ where, for instance, $S_{\varepsilon Y} = T^{-1}\sum_{t=1}^T \varepsilon_t(Y_{t-1}|D_{t-1})'$. Then bound $S_{\varepsilon Y}S_{YY}^{-1}S_{Y\varepsilon}$ using Nielsen (2005, Theorem 2.4) assuming D, E.

(ii, iii) Assuming B, E then $MY_t = (U_t, V_{\backslash 1,t}, W_t)$. These components are asymptotically uncorrelated *a.s.* given D_{t-1}, see Nielsen (2005, Theorem 9.1, 9.2, 9.4), so they can be treated individually. These terms on the one hand and on the other hand $R_C = \{(\sum_{s=1}^{t-1}\varepsilon_s', d_{t-1})'|D_{1,t-1}\}$ are asymptotically uncorrelated given $D_{\backslash 1,t-1}$. This holds *a.s.* for U_t, W_t, see Nielsen (2005, Theorem 9.2, 9.4), and in probability for $V_{\backslash 1,t}$, see Chan and Wei (1988), Chan (1989). It follows that $S_{\beta_\perp\beta_\perp\cdot\beta} = \mathcal{B}S_{CC}\mathcal{B}'\{1+\mathrm{o_P}(1)\}$.

For the models $\mathsf{M}_l, \mathsf{M}_c, \mathsf{M}_z$ then $\mathcal{B} = I_{p+1}$. Note that $R_{0\cdot\beta,t} = (\varepsilon_t|Y_{t-1}, D_{t-1})$. It then holds in a similar way $S_{0\beta_\perp\cdot\beta}S_{CC}^{-1/2} = S_{\varepsilon C}S_{CC}^{-1/2} + \mathrm{o_P}(1)$. Then apply Theorem 5.1 to get the limiting result.

For the models $\mathsf{M}_{lq}, \mathsf{M}_{cl}$ then $\mathcal{B} = \{I_{p-r}, \alpha_\perp'\mu_1(1,0)'\}$ is not a square matrix so the analysis has to take into account that d_{t-1} is of larger order than $\sum_{s=1}^{t-1}\varepsilon_s$. A rotation argument can be applied as in the proof of Johansen (1995, Theorem 11.1). ■

A.5 Analysis of the generalised common trends, not assuming B

When Assumption B does not hold the Granger-Johansen representation in Theorem 3.2 fails in that the process can be integrated of higher order. A result giving the order of $S_{0\beta_\perp\cdot\beta}S_{\beta_\perp\beta_\perp\cdot\beta}^{-1}S_{\beta_\perp 0\cdot\beta}$ can then be established using Lemma A.1 in conjunction with a more general representation result.

The general representation comes about by extending the companion form arguments of §A.3. Let $S_t = (X_t'\beta_\perp, d_t, Y_t', D_t')$ recalling $Y_t = (X_t^{*\prime}\beta^*, \Delta X_t', \ldots, \Delta X_{t-k+2}')'$ of Theorem 3.2. This vector satisfies the equation $S_t = \mathbf{S}S_{t-1} + \iota_S\varepsilon_t$ where, recalling the definitions of $\mathbf{Y}, \iota_Y, \nu$ in (A.1), it holds

$$\mathbf{S} = \begin{pmatrix} I_{p-r} & 0 & \beta_\perp'\nu & \beta_\perp'\mu \\ 0 & 1 & 0 & (1,0) \\ 0 & 0 & \mathbf{Y} & \mu_Y \\ 0 & 0 & 0 & \mathbf{D} \end{pmatrix}, \qquad \iota_S = \begin{pmatrix} \beta_\perp' \\ 0 \\ \iota_Y \\ 0 \end{pmatrix}. \tag{A.14}$$

The companion matrix can be decomposed following the argument in Nielsen (2005, §3). In general the matrix $\mathbf{Y}$ may have some unit roots. Thus, there exists a regular, deterministic matrix M and matrices m_{1Y}, m_{1D}, m_{YD} so that

$$\tilde{S}_t = \begin{pmatrix} I_{p-r} & 0 & m_{1Y} & m_{1D} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & M & m \\ 0 & 0 & 0 & I_{\dim \mathbf{D}} \end{pmatrix} S_t = \begin{pmatrix} \beta'_\perp X_t + m_{1Y} Y_t + m_{1D} D_t \\ d_t \\ (U'_t, V'_{1,t}, V'_{\backslash 1,t}, W'_t)' \\ (D'_{1,t}, D'_{\backslash 1,t})' \end{pmatrix},$$

satisfies the equation $\tilde{S}_t = \tilde{\mathbf{S}} \tilde{S}_{t-1} + \tilde{\iota}_S \varepsilon_t$ where, for some $\nu_1, \mu_{11}, \mu_{Y1}, \mu_{Y\backslash 1}$, it holds

$$\tilde{\mathbf{S}} = \left(\begin{array}{cc|cccc|cc} I_{p-r} & 0 & 0 & \nu_1 & 0 & 0 & \mu_{11} & 0 \\ 0 & 1 & 0 & \cdots & \cdots & 0 & (1,0) & \vdots \\ \hline \vdots & \ddots & \mathbf{U} & \ddots & & \vdots & & \vdots \\ \vdots & & \ddots & \mathbf{V}_1 & \ddots & \vdots & \mu_{Y1} & 0 \\ \vdots & & & \ddots & \mathbf{V}_{\backslash 1} & 0 & 0 & \mu_{Y\backslash 1} \\ \vdots & & & & \ddots & \mathbf{W} & 0 & 0 \\ \hline \vdots & & & & & \ddots & \mathbf{D}_1 & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & \mathbf{D}_{\backslash 1} \end{array}\right), \quad \tilde{\iota}_S = \left(\begin{array}{c} \beta'_\perp + m_{1Y}\iota_Y \\ 0 \\ \hline M\iota_Y \\ \hline 0 \\ 0 \end{array}\right).$$

Here $|\text{eigen}(\mathbf{U})| < 1$, $\text{eigen}(\mathbf{V}_1) = 1$, $|\text{eigen}(\mathbf{V}_{\backslash 1})| = 1$ but $\text{eigen}(\mathbf{V}_{\backslash 1}) \neq 1$, and $|\text{eigen}(\mathbf{W})| > 1$. When Assumption B does not hold the parameters $\mathbf{V}_1, \nu_1, \mu_{Y1}$ can generate higher order integrated components along with higher order deterministic polynomials. These unit root components satisfy $\breve{V}_t = \breve{\mathbf{V}} \breve{V}_{t-1} + \iota_{\breve{V}} \varepsilon_t$, where

$$\breve{V}_t = \begin{pmatrix} \beta'_\perp X_t + m_{1Y} Y_t + m_{1D} D_t \\ d_t \\ V_{1,t} \\ D_{1,t} \end{pmatrix}, \quad \breve{\mathbf{V}} = \begin{pmatrix} I_{p-r} & 0 & \nu_1 & \mu_{11} \\ 0 & 1 & 0 & (1,0) \\ \vdots & \ddots & \mathbf{V}_1 & \mu_{Y1} \\ 0 & \cdots & 0 & \mathbf{D}_1 \end{pmatrix},$$

and $\iota_{\breve{V}}$ is defined conformably from $\iota_{\tilde{S}}$. Since it is ultimately of interest to analyse the residuals of $\beta'_\perp X_{t-1}$ given Y_{t-1}, D_{t-1} the term $m_{1Y} Y_t + m_{1D} D_t$ in the first component of $\breve{V}_t$ will not play any role and is ignored in the subsequent manipulations.

The next step is to separate unit root and deterministic components as in Nielsen (2005, equation 3.5). That is, there exists a matrix $\breve{m}$ so $\breve{V}_t = \breve{m} \tilde{L}_t$ where

$$\tilde{L}_t = \begin{pmatrix} \tilde{V}_t \\ \tilde{D}_t \end{pmatrix} = \begin{pmatrix} \tilde{\mathbf{V}}_1 & 0 \\ 0 & \tilde{\mathbf{D}}_1 \end{pmatrix} \begin{pmatrix} \tilde{V}_{t-1} \\ \tilde{D}_{t-1} \end{pmatrix} + \begin{pmatrix} \iota_{\tilde{V}} \\ 0 \end{pmatrix} \varepsilon_t$$

where $\tilde{\mathbf{V}}_1$ is a $(p-r+\dim \mathbf{V}_1)$-dimensional block diagonal matrix so the diagonal blocks are Jordan blocks of the type (2.3) while $\tilde{\mathbf{D}}_1$ is a $(\dim \tilde{\mathbf{V}}_1+1+\dim \mathbf{D}_1)$-dimensional Jordan blocks of the type (2.3), and where $\iota_{\tilde{V}}$ is defined from the $(\beta'_\perp X_t, V_{1,t})$-components of $\iota_{\check{V}}$. Thus, consider the following result.

Lemma A.10 *Assuming A, D with $\gamma > 1$. Let $\tilde{K}_t = (U'_t, V'_{\backslash 1,t}, \lambda'_t, D_{\backslash 1,t})'$. Then*
(i) $S_{\tilde{L}\tilde{L}\cdot\tilde{K},w'_\perp Z} = S_{\tilde{L}\tilde{L}}\{1 + \mathrm{o_P}(1)\}$,
(ii) $S_{\varepsilon\tilde{L}\cdot\tilde{K},w'_\perp Z} S_{\tilde{L}\tilde{L}}^{-1/2} = \mathrm{O_P}(T^{-1/2})$.

Proof of Lemma A.10. (i) First, Z_{t-1} is asymptotically uncorrelated with $\tilde{K}_{t-1}, \tilde{L}_{t-1}$ in probability by Nielsen (2008, Theorem 4.7) assuming A, D with $\gamma > 1$. Secondly, it is argued that $\tilde{K}_{t-1}$ and $\tilde{L}_{t-1}$ are asymptotically uncorrelated in probability. To see this note that U_{t-1}, λ_{t-1} are asymptotically uncorrelated with each other and with the remaining terms *a.s.* by Nielsen (2005, Theorem 9.1, 9.2, 9.4) and note that $V_{\backslash 1,t-1}, D_{\backslash 1,t-1}$ are asymptotically uncorrelated with $\tilde{L}_{t-1}$ in probability by arguments as in Chan and Wei (1988), Chan (1989).

(ii) Note first that $S_{\varepsilon\tilde{L}\cdot\tilde{K},w'_\perp Z} = n_T S_{(\varepsilon,w'_\perp Z),\tilde{L}\cdot\tilde{K}}$ where $n_T = (I_p, -S_{\varepsilon,w'_\perp Z} S^{-1}_{w'_\perp Z,w'_\perp Z})$ is convergent by the Law of Large Numbers in (A.11) assuming A, D with $\gamma > 0$. Then write $S_{(\varepsilon,w'_\perp Z),\tilde{L}\cdot\tilde{K}} S_{\tilde{L}\tilde{L}}^{-1/2} = S_{(\varepsilon,w'_\perp Z),\tilde{L}} S_{\tilde{L}\tilde{L}}^{-1/2} - S_{(\varepsilon,w'_\perp Z),\tilde{K}} S^{-1}_{\tilde{K}\tilde{K}} S_{\tilde{K}\tilde{L}} S_{\tilde{L}\tilde{L}}^{-1/2}$.

The first term $S_{(\varepsilon,w'_\perp Z),\tilde{L}} S_{\tilde{L}\tilde{L}}^{-1/2}$: note that $T^{-1/2}\sum_{s=1}^{\mathrm{int}(Tu)} (\varepsilon'_s, Z'_{s-1}w_\perp)'$ is asymptotically Brownian assuming D, see (5.3). The vector $\tilde{L}_t$, which has unit root and polynomial components is a continuous function of $\sum_{s=1}^{\mathrm{int}(Tu)} \varepsilon_s$ and there exists a normalisation matrix $\mathcal{N}_{\tilde{L},T}$ so $\mathcal{N}_{\tilde{L},T}\tilde{L}_{\mathrm{int}(Tu)}$ has a non-degenerate limit, see (5.4). Theorem 5.1 then implies $S_{(\varepsilon,w'_\perp Z),\tilde{L}} S_{\tilde{L}\tilde{L}}^{-1/2} = \mathrm{O_P}(T^{-1/2})$.

The second term, $S_{(\varepsilon,w'_\perp Z),\tilde{K}} S_{\tilde{K}\tilde{K}}^{-1/2}$ is $\mathrm{O}(1)$ *a.s.* due to the Law of Large Numbers in (A.11) and since the correlation matrix $S^{-1/2}_{(\varepsilon,w'_\perp Z),(\varepsilon,w'_\perp Z)} S_{(\varepsilon,w'_\perp Z),\tilde{K}} S_{\tilde{K}\tilde{K}}^{-1/2}$ is $\mathrm{O}(1)$.

The third term $S_{\tilde{K}\tilde{K}}^{-1/2} S_{\tilde{K}\tilde{L}} S_{\tilde{L}\tilde{L}}^{-1/2}$ is $\mathrm{O_P}(T^{-1/2})$. To see this apply an argument as in Chan and Wei (1988) and Chan (1989). ■

Lemma A.11 *Assuming A, D with $\gamma > 1$ then $S_{0\beta_\perp\cdot\beta} S^{-1}_{\beta_\perp\beta_\perp\cdot\beta} S_{\beta_\perp 0\cdot\beta} = \mathrm{O_P}(T^{-1})$.*

Proof of Lemma A.11. Note that $R_{0\cdot\beta,t} = (\varepsilon_t | Y_{t-1}, D_{t-1})$. So it suffices to show $S_{\varepsilon\beta_\perp\cdot\beta} S^{-1}_{\beta_\perp\beta_\perp\cdot\beta} S_{\beta_\perp\varepsilon\cdot\beta} = \mathrm{O_P}(T^{-1})$. Recall the definitions of $\tilde{K}_t, \tilde{L}_t$. Then Lemma A.1 shows $S_{\varepsilon\beta_\perp\cdot\beta} S^{-1}_{\beta_\perp\beta_\perp\cdot\beta} S_{\beta_\perp\varepsilon\cdot\beta} = \mathrm{O}(S_{\varepsilon\tilde{L}\cdot w'_\perp Z,\tilde{K}} S^{-1}_{\tilde{L}\tilde{L}\cdot w'_\perp Z,\tilde{K}} S_{\tilde{L}\varepsilon\cdot w'_\perp Z,\tilde{K}})$. The latter matrix is $\mathrm{O_P}(T^{-1})$ by Lemma A.10(i, ii). ■

A.6 Proof of consistency

Proof of Theorem 4.1. Recall the rewritten eigenvalue problem (A.7), that is

$$0 = \det\{P(\lambda)\}, \quad P(\lambda) = \lambda S_{00\cdot\beta} + (\lambda - 1) S_{0\beta} S_{\beta\beta}^{-1} S_{\beta 0} - S_{0\beta_\perp\cdot\beta} S_{\beta_\perp\beta_\perp\cdot\beta}^{-1} S_{\beta_\perp 0\cdot\beta}. \quad \text{(A.15)}$$

Note that $\text{rank}(S_{0\beta} S_{\beta\beta}^{-1} S_{\beta 0}) \leq r$ and $\text{rank}(S_{0\beta_\perp\cdot\beta} S_{\beta_\perp\beta_\perp\cdot\beta}^{-1} S_{\beta_\perp 0\cdot\beta}) \leq p - r$ indicating how the eigenvalues can be separated. For the weak consistency result it suffices

(a) If ρ_r is the smallest non-zero eigenvalue of $S_{0\beta} S_{\beta\beta}^{-1} S_{\beta 0}$ then $\rho_r^{-1} = \mathrm{O}_\mathsf{P}(1)$.

(b) The limit of $S_{00\cdot\beta}$ has full rank.

(c) $S_{0\beta_\perp\cdot\beta} S_{\beta_\perp\beta_\perp\cdot\beta}^{-1} S_{\beta_\perp 0\cdot\beta} = \mathrm{O}_\mathsf{P}(T^{-1})$.

For result (a) note first that by Lemma A.4(ii) assuming A, D with $\gamma > 1$ then

$$A = \overline{\alpha}_{\lim}' S_{0\beta} S_{\beta\beta}^{-1} S_{\beta 0} \overline{\alpha}_{\lim} = \{I_r + \mathrm{o}_\mathsf{P}(1)\} S_{\beta\beta} \{I_r + \mathrm{o}_\mathsf{P}(1)\}.$$

Since $S_{\beta\beta}^{-1} = \mathrm{O}_\mathsf{P}(1)$ by Lemma A.4(iii) assuming A, D with $\gamma > 1$ then the smallest eigenvalue, $\tilde{\rho}_r$ say, of A satisfies $\tilde{\rho}_r^{-1} = \mathrm{O}_\mathsf{P}(1)$. Since $\tilde{\rho}_r \leq \rho_r$ by Poincaré's separation theorem, see Magnus and Neudecker (1988, Theorem 11.12), then also $\rho_r^{-1} = \mathrm{O}_\mathsf{P}(1)$.

Here (b) follows from Lemma A.4(i) while (c) follows from Lemma A.11, both assuming A, D with $\gamma > 1$. ∎

Proof of Theorem 4.2. Follow the proof of Theorem 4.1 with two modifications. Apply Lemma A.5 assuming A, D, E with $\gamma > 0$ instead of Lemma A.4. Apply Lemma A.7 assuming A, D, E with $\gamma > 0$ instead of Lemma A.11. Note that different rates apply depending on whether Assumption C holds or not. ∎

A.7 Proof of asymptotic distribution of rank test

Proof of Theorem 4.4. The solutions to the eigenvalue problem (A.15) equal to those of $0 = \det\{\tilde{P}(\varrho)\}$ where

$$\tilde{P}(\varrho) = A_T' P(T^{-1}\varrho) A_T = \left\{ \begin{array}{cc} P_{\alpha\alpha}(T^{-1}\varrho) & P_{\alpha\alpha_\perp}(T^{-1}\varrho) \\ P_{\alpha_\perp\alpha}(T^{-1}\varrho) & P_{\alpha_\perp\alpha_\perp}(T^{-1}\varrho) \end{array} \right\},$$

with $A_T = (A_{\alpha,T}, A_{\alpha_\perp,T}) = (\overline{\alpha} S_{\beta\beta}^{-1/2}, \tilde{\alpha}_\perp T^{1/2})$ and $\tilde{\alpha}_\perp = \alpha_\perp (\alpha_\perp' \Omega \alpha_\perp)^{-1/2}$. To describe $\tilde{P}(\varrho)$ note that Lemmas A.4, A.9 assuming A, B, D, E with $\gamma > 0$ show

$$\begin{aligned} S_{00\cdot\beta} &= \Omega + \mathrm{o}_\mathsf{P}(1), \quad S_{\beta\beta}^{-1} = \mathrm{O}_\mathsf{P}(1), \quad S_{0\beta_\perp\cdot\beta} S_{\beta_\perp\beta_\perp\cdot\beta}^{-1} S_{\beta_\perp 0\cdot\beta} = \mathrm{O}_\mathsf{P}(T^{-1}), \\ S_{0\beta} S_{\beta\beta}^{-1} &= \alpha + S_{\varepsilon\beta} S_{\beta\beta}^{-1}, \quad S_{\varepsilon\beta} S_{\beta\beta}^{-1} S_{\beta\varepsilon} = \mathrm{o}_\mathsf{P}(T^{-\xi}) \quad \text{for all } \xi < \gamma/(2+\gamma). \end{aligned}$$

Moreover, $T \tilde{\alpha}_\perp' S_{0\beta_\perp\cdot\beta} S_{\beta_\perp\beta_\perp\cdot\beta}^{-1} S_{\beta_\perp 0\cdot\beta} \tilde{\alpha}_\perp \xrightarrow{\mathsf{D}} \mathbf{B} = \int_0^1 dB_u F_u' (\int_0^1 F_u F_u' du)^{-1} \int_0^1 F_u dB_u'$.

In the following results for the components of $\tilde{P}(\varrho)$ are given. For the term $A'_{\alpha,T}S_{0\beta}S_{\beta\beta}^{-1}S_{\beta 0}A_{\alpha_\perp,T}$ it is needed that $S_{\varepsilon\beta}S_{\beta\beta}^{-1}S_{\beta\varepsilon} = \mathsf{o}_\mathsf{P}(T^{-1/2})$ which holds if $\gamma > 2$.

$$\begin{aligned}
A'_{\alpha,T}S_{00\cdot\beta}A_{\alpha,T} &= S_{\beta\beta}^{-1/2}\overline{\alpha}'\Omega\overline{\alpha}S_{\beta\beta}^{-1/2}\{1+\mathsf{o}_\mathsf{P}(1)\}, \\
A'_{\alpha,T}S_{00\cdot\beta}A_{\alpha_\perp,T} &= T^{1/2}S_{\beta\beta}^{-1/2}\overline{\alpha}'\Omega\tilde{\alpha}_\perp\{1+\mathsf{o}_\mathsf{P}(1)\}, \\
A'_{\alpha_\perp,T}S_{00\cdot\beta}A_{\alpha_\perp,T} &= TI_{p-r}\{1+\mathsf{o}_\mathsf{P}(1)\}, \\
A'_{\alpha,T}S_{0\beta}S_{\beta\beta}^{-1}S_{\beta 0}A_{\alpha,T} &= I_r + \mathsf{o}_\mathsf{P}(1), \\
A'_{\alpha,T}S_{0\beta}S_{\beta\beta}^{-1}S_{\beta 0}A_{\alpha_\perp,T} &= T^{1/2}S_{\beta\beta}^{-1/2}S_{\beta\varepsilon}\tilde{\alpha}_\perp + \mathsf{o}_\mathsf{P}(1). \\
A'_{\alpha_\perp,T}S_{0\beta}S_{\beta\beta}^{-1}S_{\beta 0}A_{\alpha_\perp,T} &= T\tilde{\alpha}'_\perp S_{\varepsilon\beta}S_{\beta\beta}^{-1}S_{\beta\varepsilon}\tilde{\alpha}_\perp, \\
A'_{\alpha,T}S_{0\beta_\perp\cdot\beta}S_{\beta_\perp\beta_\perp\cdot\beta}^{-1}S_{\beta_\perp 0\cdot\beta}A_{\alpha,T} &= \mathsf{o}_\mathsf{P}(1), \\
A'_{\alpha,T}S_{0\beta_\perp\cdot\beta}S_{\beta_\perp\beta_\perp\cdot\beta}^{-1}S_{\beta_\perp 0\cdot\beta}A_{\alpha_\perp,T} &= \mathsf{o}_\mathsf{P}(1), \\
A'_{\alpha_\perp,T}S_{0\beta_\perp\cdot\beta}S_{\beta_\perp\beta_\perp\cdot\beta}^{-1}S_{\beta_\perp 0\cdot\beta}A_{\alpha_\perp,T} &= \mathbf{B} + \mathsf{o}_\mathsf{P}(1).
\end{aligned}$$

It follows that

$$\begin{aligned}
P_{\alpha\alpha}(T^{-1}\varrho) &= -I_r + \mathsf{o}_\mathsf{P}(1), \\
P_{\alpha\alpha_\perp}(T^{-1}\varrho) &= -T^{1/2}S_{\beta\beta}^{-1/2}S_{\beta\varepsilon}\tilde{\alpha}_\perp + \mathsf{o}_\mathsf{P}(1), \\
P_{\alpha_\perp\alpha_\perp}(T^{-1}\varrho) &= \varrho I_{p-r} - \mathbf{B} + \mathsf{o}_\mathsf{P}(1) - T\tilde{\alpha}'_\perp S_{\varepsilon\beta}S_{\beta\beta}^{-1}S_{\beta\varepsilon}\tilde{\alpha}_\perp.
\end{aligned}$$

By the partitioned inversion formula the eigenvalue problem is rewritten as

$$0 = \det\{\tilde{P}(\varrho)\} = \det\{P_{\alpha\alpha}(T^{-1}\varrho)\}\det\{P_{\alpha_\perp\alpha_\perp\cdot\alpha}(T^{-1}\varrho)\},$$

where $P_{\alpha_\perp\alpha_\perp\cdot\alpha} = P_{\alpha_\perp\alpha_\perp} - P_{\alpha_\perp\alpha}P_{\alpha\alpha}P_{\alpha\alpha_\perp}$. Inserting the above results gives

$$0 = \det\{\tilde{P}(\varrho)\} = \det\{-I_r + \mathsf{o}_\mathsf{P}(1)\}\det\{\varrho I_{p-r} - \mathbf{B} + \mathsf{o}_\mathsf{P}(1)\}.$$

The eigenvalues of the second matrix have the desired trace. ■

Remark A.12 *In the proof of Theorem 4.4 it is used that $\gamma > 2$ as opposed to $\gamma > 0$ to ensure that $S_{\varepsilon\beta}S_{\beta\beta}^{-1}S_{\beta\varepsilon} = \mathsf{o}_\mathsf{P}(T^{-1/2})$. For non-explosive cases that result could be proved along the lines of Chan and Wei (1988) and Chan (1989) assuming $\gamma > 0$.*

A.8 On the limit distribution for singular cases

Remark 4.7 gives an example of singular explosive process. The Granger-Johansen representation for this process is

$$X_t = \frac{1}{1-\rho}\sum_{s=1}^{t}\varepsilon_t - \rho\Delta X_t = \frac{1}{1-\rho}\sum_{s=1}^{t}\varepsilon_t + \frac{1}{\rho-1}\rho^t\sum_{s=1}^{t}\rho^{-s}\varepsilon_t;$$

see Theorem 3.2 or Nielsen (2008, Theorem 1). As a consequence

$$\begin{aligned} R_{0,t} &= (\Delta X_t \mid \Delta X_{t-1}) = (\varepsilon_t \mid \Delta X_{t-1}), \\ R_{1,t} &= (X_{t-1} \mid \Delta X_{t-1}) = (\frac{1}{1-\rho}\sum_{s=1}^{t}\varepsilon_t \mid \Delta X_{t-1}). \end{aligned}$$

Moreover, as in (A.9) it holds that

$$\Delta X_t = \rho^t \sum_{s=1}^{t} \rho^{-s}\varepsilon_s = w\lambda_t - Z_t, \tag{A.16}$$

where $\lambda_t = \rho^t$, $Z_t = \sum_{s=1}^{\infty} \rho^{-s}\varepsilon_{t+j}$ and $w = Z_0$. Thus, in $R_{0,t}, R_{1,t}$ the regressor ΔX_{t-1} can be replaced by $\rho^{t-1}, w'_\perp Z_{t-1}$. Due to the uncorrelatedness of $\varepsilon_t, \sum_{s=1}^{t}\varepsilon_s$ with ρ^{t-1} and the uncorrelatedness of $\sum_{s=1}^{t}\varepsilon_s$ with $w'_\perp Z_{t-1}$ then

$$\begin{aligned} (1-\rho)^2 T^{-1} S_{11} &= T^{-2}\sum_{t=1}^{T}(\sum_{s=1}^{t-1}\varepsilon_s)^{\otimes 2}\{1+\mathsf{o}_\mathsf{P}(1)\}, \\ (1-\rho)^2 S_{10} &= T^{-1}\sum_{t=1}^{T}(\sum_{s=1}^{t-1}\varepsilon_s)(\varepsilon_t \mid w'_\perp Z_{t-1})' + \mathsf{o}_\mathsf{P}(1), \\ S_{00} &= T^{-1}\sum_{t=1}^{T}(\varepsilon_t \mid w'_\perp Z_{t-1})^{\otimes 2}. \end{aligned}$$

By Theorem 5.1 it then holds

$$(1-\rho)^2 T^{-1} S_{11} \overset{\mathsf{D}}{\to} \mathcal{I}_{11}(I-aP), \quad (1-\rho)^2 S_{10} \overset{\mathsf{D}}{\to} \mathcal{I}_{10}(I-aP),$$

and $S_{00} \to I - aP$ in probability, where

$$\mathcal{I}_{11} = \textstyle\int_0^1 B_u B'_u du \quad \mathcal{I}_{10} = \int_0^1 B_u dB'_u, \quad P = w_\perp(w'_\perp w_\perp)^{-1}w'_\perp, \quad a = \dfrac{\rho^2-1}{\rho^2}.$$

It follows that

$$\mathsf{LR} = \mathrm{tr}(S_{00}^{-1}S_{01}S_{11}^{-1}S_{10}) + \mathsf{o}_\mathsf{P}(1) = \mathrm{tr}\{(I-a^{-1}P)\mathcal{I}_{01}\mathcal{I}_{11}^{-1}\mathcal{I}_{10}\} + \mathsf{o}_\mathsf{P}(1).$$

A.9 Stochastic integrals

Proof of Theorem 5.1. It is argued that the second term in (5.5) vanishes. Apply partial summation formula, with $Z_{T,t} = T^{-1/2}Z_t$ and $H_{T,t-1} = h(J_T)_{(t-1)/T}$, to get

$$\sum_{t=1}^{\mathrm{int}(Tu)} (H_{T,t-1} - H_{T,0})\Delta Z'_{T,t} = (H_{T,\mathrm{int}(Tu)} - H_{T,0})Z'_{T,\mathrm{int}(Tu)} - \sum_{t=1}^{\mathrm{int}(Tu)} \Delta H_{T,t} Z'_{T,t}.$$

First term. Note that $H_{T,\text{int}(Tu)}$ is convergent to a continuous process so its supremum also converges, while $Z_{T,\text{int}(Tu)} = \text{o}(1)$ *a.s.* uniformly in u by Nielsen (2008, Corollary 4.3). Thus the first term vanishes.

Second term. Note that $Z_{T,t} = T^{-1/2}\sum_{j=1}^{\infty}\mathbf{W}^{-j}e_{W,t+j}$ and rewrite the second term $\sum_{t=1}^{\text{int}(Tu)}\Delta H_{T,t}Z'_{T,t} = \mathcal{I}_{1,T,u} + \mathcal{I}_{2,T,u}$ where

$$\begin{aligned}
\mathcal{I}_{1,T,u} &= T^{-1/2}\sum_{s=2}^{\text{int}(Tu)}\sum_{t=1}^{s-1}\Delta H_{T,t}(\mathbf{W}^{t-s}e_{W,s})', \\
\mathcal{I}_{2,T,u} &= T^{-1/2}\sum_{t=1}^{\text{int}(Tu)}\Delta H_{T,t}\sum_{s=1}^{\infty}(\mathbf{W}^{-T+t-s}e_{W,T+s})'.
\end{aligned}$$

The term $\mathcal{I}_{1,T,u}$. It suffices to consider each coordinate of $\mathbf{W}^{t-s}e_{W,s}$. Assume this is univariate or apply a Jordan decomposition argument. Since $e_{W,s}$ is a martingale difference and $\Delta H_{T,t}$ is $\mathcal{F}_{s-1}$-adapted then by Lai and Wei (1982, Lemma 1)

$$\mathcal{I}_{1,T,u} = T^{-1/2}\text{O}[\log\lambda_{\max}\{\sum_{s=2}^{\text{int}(Tu)}(\sum_{t=1}^{s-1}\Delta H_{T,t}\mathbf{W}^{t-s})^{\otimes 2}] \quad a.s.$$

It suffices to argue that the double sum is of polynomial order. By the Cauchy-Schwarz inequality and noting $|\text{eigen}(\mathbf{W}^{-1})| \leq 1$ then

$$\begin{aligned}
\sum_{s=2}^{\text{int}(Tu)}(\sum_{t=1}^{s-1}\Delta H_{T,t}\mathbf{W}^{t-s})^{\otimes 2} &\leq \sum_{s=2}^{\text{int}(Tu)}(\sum_{t=1}^{s-1}||\Delta H_{T,t}||^2)(\sum_{t=1}^{s-1}||\mathbf{W}||^{2(t-s)}) \\
&\leq (\sum_{s=2}^{\text{int}(Tu)}\sum_{t=1}^{s-1}||\Delta H_{T,t}||^2)(\sum_{t=1}^{\infty}||\mathbf{W}||^{-2t}).
\end{aligned}$$

The mean value theorem gives

$$||\Delta H_{T,t}|| = ||\{\frac{\partial}{\partial u}h(\cdot,\cdot)(u)\Big|_{(J_T,\tau),u=t/T}\}\{T^{-1/2}(\varepsilon_t^{*\prime},u^*)'\}||$$

for some ε_t^* and u^* so $||\varepsilon_t^*|| \leq ||\varepsilon_t||$ and $0 \leq u^* \leq u^{1/2} \leq 1$. This is bounded by

$$||\Delta H_{T,t}|| \leq \{\sup_{0\leq u\leq 1}||\frac{\partial}{\partial u}h(\cdot,\cdot)(u)\Big|_{(J_T,\tau),u=t/T}||\}\{T^{-1/2}\sup_{t\leq T}(||\varepsilon_t||,1)\}$$

Due to the assumed bound to h and since $||\varepsilon_t||^2 = \text{o}(T^{1-\xi})$ for all $\xi < \gamma/(2+\gamma)$, see Lai and Wei (1985, Theorem 1) then $||\Delta H_{T,t}|| = \text{o}(T^{-\xi/2})$ uniformly in t. Thus, the above double sum is of polynomial order in T, uniformly in u with probability one. In turn $||\mathcal{I}_{1,T,u}|| = \text{o}(1)$ *a.s.* uniformly in u.

The term $\mathcal{I}_{2,T,u}$. Apply first the triangle inequality to get a uniform bound in u

$$||\mathcal{I}_{2,T,u}|| \leq \mathcal{J}_T = (\sum_{t=1}^{T} ||\mathbf{W}^{-1}||^{T-t}||T^{-1/2}\Delta H_{T,t}||)(||\sum_{s=1}^{\infty} \mathbf{W}^{-s} e_{W,T+s}||).$$

It holds $\mathcal{J}_T = \mathrm{o}(1)$ $a.s.$ if for all constants $K > 0$ it holds $\sum_{T=1}^{\infty} 1(||\mathcal{J}_T|| > K) = \sum_{T=1}^{\infty} 1(||\mathcal{J}_T||^{\alpha} > K^{\alpha}) < \infty$ $a.s.$ for any $\alpha > 0$. By the conditional Borel-Cantelli lemma of Chen (1978) this holds $a.s.$ on the set where $\sum_{T=1}^{\infty} \mathsf{P}(||\mathcal{J}_T||^{\alpha} > K^{\alpha}|\mathcal{F}_T) < \infty$.

Now, by the Markov inequality

$$\mathsf{P}(||\mathcal{J}_T||^{\alpha} > K^{\alpha}|\mathcal{F}_T) \leq \frac{1}{K^{\alpha}}\mathsf{E}(||\mathcal{J}_T||^{\alpha}|\mathcal{F}_T), \tag{A.17}$$

so it suffices to show $\mathsf{E}(||\mathcal{J}_T||^{\alpha}|\mathcal{F}_T) = \mathrm{o}(T^{-\zeta})$ for some $\zeta > 1$.

The expectation $\mathsf{E}(||\mathcal{J}_T||^{\alpha})$ may be undefined. In that case apply the truncation argument in the proof of Lai and Wei (1982, Lemma 2): Choose constants a_t so $\mathsf{P}(||\Delta H_{T,t}||^{\alpha} > a_t) < t^{-2}$. By the Borel-Cantelli Lemma, see Breiman (1968, p.41), then $\mathsf{P}(\Delta H_{T,t} = \Delta H^*_{T,t}$ for large $t) = 1$ where $\Delta H^*_{T,t} = \Delta H_{T,t}$ if $||\Delta H_{T,t}||^{\alpha} < a_t$ and zero otherwise.

To bound $\mathsf{E}(||\mathcal{J}_T||^{\alpha}|\mathcal{F}_T)$ note that a sum $n_t = \sum_{j=1}^{\infty} a_j m_j$ can be bounded using the spectral norm inequality $||a_j m_{t+j}|| \leq ||a_j||\,||m_{t+j}||$ and the Jensen inequality through the inequality $||n_t||^{\alpha} \leq (\sum_{j=1}^{\infty} ||a_j||)^{\alpha-1} \sum_{j=1}^{\infty} ||a_j||||m_{t+j}||^{\alpha}$ for $\alpha > 1$, see also Nielsen (2008, equation 4.2). Apply this bound to each of the sums in $\mathcal{J}_T$ to get

$$||\mathcal{J}_T||^{\alpha} \leq c_T(\sum_{t=1}^{T} ||\mathbf{W}^{-1}||^{T-t}||T^{-1/2}\Delta H_{T,t}||^{\alpha})(\sum_{s=1}^{\infty} ||\mathbf{W}^{-1}||^{s}||e_{W,T+s}||^{\alpha}),$$

where $c_T = ||\mathbf{W}^{-1}||^{\alpha-1}(1 - ||\mathbf{W}^{-1}||^{T})^{\alpha-1}(1 - ||\mathbf{W}^{-1}||)^{2(1-\alpha)}$ is bounded uniformly in T. Noting that $\Delta H_{T,t}$ is $\mathcal{F}_T$-measurable then

$$\mathsf{E}(||\mathcal{J}_T||^{\alpha}|\mathcal{F}_T) \leq c_T(\sum_{t=1}^{T} ||\mathbf{W}||^{t-T}||T^{-1/2}\Delta H_{T,t}||^{\alpha})\{\sum_{s=1}^{\infty} ||\mathbf{W}||^{-s}\mathsf{E}(||e_{W,T+s}||^{\alpha}|\mathcal{F}_T)\}.$$

By Assumption D then $\sup_t \mathsf{E}(||e_{W,T+s}||^{\alpha}|\mathcal{F}_T) < \infty$ $a.s.$ for $\alpha < 2+\gamma$, which implies that the sum in s is finite $a.s.$ For the sum in t use the bound from above that $||\Delta H_{T,t}|| = \mathrm{o}(T^{-\xi/2})$, uniformly in t. Thus, the sum in t is bounded by the product of $\mathrm{o}(T^{-(1+\xi)\alpha/2})$ and the bounded sum $\sum_{t=1}^{T} ||\mathbf{W}||^{t-T}$. Thus, the sum in t is $\mathrm{o}(T^{-(1+\xi)\alpha/2})$. Since it must hold that $(1+\xi)\alpha/2 > 1$ while $\xi < \gamma/(2+\gamma)$ and $\alpha < 2+\gamma$ then $(1+\xi)\alpha/2 < \{2(1+\gamma)/(2+\gamma)\}(2+\gamma)/2 \leq 1+\gamma$. Thus $(1+\xi)\alpha/2$ can be chosen larger than unity for any $\gamma > 0$. ■

References

Anderson, T.W. (1959) On asymptotic distributions of estimates of parameters of stochastic difference equations. *Annals of Mathematical Statistics* 30, 676-687.

Bauer, D. (2009) Almost sure bounds on the estimation error for OLS estimators when the regressors include certain MFI(1) processes. *Econometric Theory* 25, 571-582.

Breiman, L. (1968) *Probability.* Reading, MA: Addison-Wesley.

Chan, N.H. and C.Z. Wei (1988) Limiting distributions of least squares estimates of unstable autoregressive processes. *Annals of Statistics* 16, 367-401.

Chan, N.H. (1989) Asymptotic inference for unstable autoregressive time series with drifts. *Journal of Statistical Planning and Inference* 23, 310-312.

Chen, L.H.Y. (1978) A short note on the conditional Borel-Cantelli lemma. *Annals of Probability* 6, 699-700.

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

Duflo, M., R. Senoussi & R. Touati (1991) Propriétés asymptotiques presque sûre de l'estimateur des moindres carrés d'un modèle autorégressif vectoriel. *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques* 27, 1-25.

de Jong R.M. and J. Davidson (2000) The functional central limit theorem and weak convergence to stochastic integrals I. *Econometric Theory* 16, 621-642.

Engler, E. and Nielsen, B. (2009) The empirical process of autoregressive residuals. *Econometrics Journal* 12, 367-381.

Jakubowski, A., Mémin, J. and Pages, G. (1989) Convergence en loi des suites d'intégrales stochastiques sur l'espace D^1 de Skorokhod. *Probability Theory and Related Fields* 81, 111-137.

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

Johansen, S. (1995a) *Likelihood-based inference in cointegrated vector autoregressive models.* Oxford: Oxford University Press.

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

Johansen, S. and Schaumburg, E. (1999) Likelihood analysis of seasonal cointegration. *Journal of Econometrics* 88, 301-339.

Lai, T.L. & C.Z. Wei (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. & C.Z. Wei (1983) A note on martingale difference sequences satisfying the local Marcinkiewicz-Zygmund condition. *Bulletin of the Institute of Mathematics, Academia Sinica* 11, 1-13.

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

Magnus, J.R. and Neudecker, H. (1988) *Matrix differential calculus with applications in statistics and econometrics.* New York: Wiley.

Nielsen, B. (2001) The asymptotic distribution of unit root tests of unstable autoregressive processes. *Econometrica* 69, 211-219.

Nielsen, B. (2004) On the distribution of tests of cointegration. *Econometric Reviews* 23, 1-23.

Nielsen, B. (2005) Analysis of co-explosive processes. Discussion paper, Nuffield College. To appear in *Econometric Theory.*

Nielsen, B. (2006a) Order determination in general vector autoregressions. In Ho, H.-C., Ing, C.-K., and Lai, T.L. (eds): *Time Series and Related Topics: In Memory of Ching-Zong Wei.* IMS Lecture Notes and Monograph Series, 52, 93-112. doi:10.1214/074921706000000978.

Nielsen, B. (2006b) Correlograms for non-stationary autoregressions. *Journal of the Royal Statistical Society*, B68, 707-720.

Nielsen, B. (2008) Singular vector autoregressions with deterministic terms: Strong consistency and lag order determination. Discussion paper, Nuffield College.

Phillips, P.C.B. and Magdalinos, T. (2008) Limit theory for explosively cointegrated systems. *Econometric Theory* 24, 865-887.

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